

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:12:45 ; Search time 36 Seconds

(without alignments)

636.642 Million cell updates/sec

Title: US-09-674-779b-2

Sequence: 1 MMLHQTAAAAALSVLFEM.....IAFGTNEAMSNRRALSLV 172

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	172	100.0	172	21	AAV55089
2	107	62.2	172	21	AAV55090
3	107	62.2	172	21	AAV55091
4	68	39.5	172	21	AAV55092
5	15	8.7	16	21	AAV55093
6	15	8.7	16	21	AAV55094
7	14	8.1	14	21	AAV55094
8	14	8.1	14	21	AAV55094
9	9	5.2	2168	22	ABR64563
10	8	4.7	139	22	ABG22196

11	8	4.7	158	21	AAI18977	Zea mays protein f
12	8	4.7	306	22	ABR62221	Drosophila melanog
13	8	4.7	436	22	ABG11084	Novel human diagno
14	8	4.7	452	21	AAV74388	Neisseria meningit
15	8	4.7	545	22	ABR58578	Drosophila melanog
16	8	4.7	545	22	ABR67264	Drosophila melanog
17	8	4.7	606	23	AAU75084	Ryegrass 4-coumar
18	8	4.7	632	22	ABR70398	Drosophila melanog
19	8	4.7	632	22	ABR66879	Drosophila melanog
20	8	4.7	644	23	ABR14482	Rice acetolactica
21	8	4.7	644	23	AAU10024	Rice acetolactica
22	8	4.7	662	22	ABR59430	Rice herbicide res
23	8	4.7	1164	22	ABR57802	Drosophila melanog
24	8	4.7	1412	22	ABR60943	Drosophila melanog
25	8	4.7	2090	22	ABR64682	Drosophila melanog
26	8	4.7	2703	22	ABR63299	Drosophila melanog
27	8	4.7	9	18	AAW04637	Null peptide that
28	9	4.1	10	19	AAW04637	Biotinylated pepti
29	9	4.1	13	16	AAW04637	Murine MHC class I
30	9	4.1	13	16	AAW04637	AKAP79 A37-50 muta
31	9	4.1	14	17	AAW02569	Mutant peptide AKA
32	9	4.1	14	21	AAW14909	Murine MHC class I
33	9	4.1	16	16	AAW44047	Human peptide #593
34	9	4.1	21	22	ABR27942	Human brain expres
35	9	4.1	21	22	ABR37770	Human bone marrow
36	9	4.1	21	22	AAW53912	Peptide #5142 enco
37	9	4.1	21	22	AAW70864	Peptide #614 enco
38	9	4.1	21	22	AAW18708	Human peptide enco
39	9	4.1	21	22	AAW26577	Human RIZ alternat
40	9	4.1	21	22	AAW26577	5' terminus of hum
41	9	4.1	21	22	AAW26577	Human Rb-interacti
42	9	4.1	21	22	AAW26577	HIV tat related am
43	9	4.1	21	22	AAW26577	Arctic fish antifr
44	9	4.1	21	22	AAW26577	Antifreeze protein
45	9	4.1	21	22	AAW26577	Protein derived fr
46	9	4.1	21	22	AAW26577	P. americanus anti
47	9	4.1	21	22	AAW26577	P. americanus anti
48	9	4.1	21	22	AAW26577	Human Rb-interacti
49	9	4.1	21	22	AAW26577	Human Rb-interacti
50	9	4.1	21	22	AAW26577	Human Rb-interacti
51	9	4.1	21	22	AAW26577	Human Rb-interacti
52	9	4.1	21	22	AAW26577	Human Rb-interacti
53	9	4.1	21	22	AAW26577	Human Rb-interacti
54	9	4.1	21	22	AAW26577	Human Rb-interacti
55	9	4.1	21	22	AAW26577	Human Rb-interacti
56	9	4.1	21	22	AAW26577	Human Rb-interacti
57	9	4.1	21	22	AAW26577	Human Rb-interacti
58	9	4.1	21	22	AAW26577	Human Rb-interacti
59	9	4.1	21	22	AAW26577	Human Rb-interacti
60	9	4.1	21	22	AAW26577	Human Rb-interacti
61	9	4.1	21	22	AAW26577	Human Rb-interacti
62	9	4.1	21	22	AAW26577	Human Rb-interacti
63	9	4.1	21	22	AAW26577	Human Rb-interacti
64	9	4.1	21	22	AAW26577	Human Rb-interacti
65	9	4.1	21	22	AAW26577	Human Rb-interacti
66	9	4.1	21	22	AAW26577	Human Rb-interacti
67	9	4.1	21	22	AAW26577	Human Rb-interacti
68	9	4.1	21	22	AAW26577	Human Rb-interacti
69	9	4.1	21	22	AAW26577	Human Rb-interacti
70	9	4.1	21	22	AAW26577	Human Rb-interacti
71	9	4.1	21	22	AAW26577	Human Rb-interacti
72	9	4.1	21	22	AAW26577	Human Rb-interacti
73	9	4.1	21	22	AAW26577	Human Rb-interacti
74	9	4.1	21	22	AAW26577	Human Rb-interacti
75	9	4.1	21	22	AAW26577	Human Rb-interacti
76	9	4.1	21	22	AAW26577	Human Rb-interacti
77	9	4.1	21	22	AAW26577	Human Rb-interacti
78	9	4.1	21	22	AAW26577	Human Rb-interacti
79	9	4.1	21	22	AAW26577	Human Rb-interacti
80	9	4.1	21	22	AAW26577	Human Rb-interacti
81	9	4.1	21	22	AAW26577	Human Rb-interacti
82	9	4.1	21	22	AAW26577	Human Rb-interacti
83	9	4.1	21	22	AAW26577	Human Rb-interacti

84 4.1 75 22 AAC62224 Human gene 24-enco
85 4.1 75 23 ABG63540 Human albumin fusi
86 4.1 75 23 ABG63541 Human albumin fusi
87 4.1 78 22 ABG63541 Human albumin fusi
88 4.1 79 22 AAU86705 Novel human connec
89 4.1 82 20 AAU86705 Winter flounder an
90 4.1 88 21 AAU87242 Human signal pepti
91 4.1 97 20 AAU87242 P. americanus anti
92 4.1 102 21 AAU87242 Zea mays protein f
93 4.1 106 22 ABG03609 Novel human diagno
94 4.1 109 22 AAU87242 Cysteine protease
95 4.1 119 22 AAU87242 Human reproductive
96 4.1 121 23 AAU87242 Chicken anemia vir
97 4.1 125 20 AAU87242 A hypothesis of mat
98 4.1 131 20 AAU87242 A hypothesis of mat
99 4.1 131 20 AAU87242 A hypothesis of mat
100 4.1 131 21 AAU87242 A hypothesis of mat

ALIGNMENTS

RESULT 1

AAV55089 standard; Protein: 172 AA.

AAV55089;

01-MAR-2000 (first entry)

M. catarrhalis BASB019 protein sequence #1.

BASB019 protein: diagnosis; infectious organism; auditive nerve damage;
genetic mutation screening; antibody production; vaccine; otitis media;
bacterial infection; pneumonia; sinusitis; nosocomial infection;
invasive disease; delayed speech learning; bacteria adhesion prevention;
upper respiratory tract infection; middle ear infection; therapy.

Moraxella catarrhalis.

WO9957277-A2.

11-NOV-1999.

03-MAY-1999; 99WO-EP03038.

06-MAY-1998; 98GB-0009683.

(SMIRK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-062148/05.

N-PSDB: AAZ40351.

Novel BASB019 polynucleotides and polypeptides from Moraxella

catarrhalis used to prepare vaccines against bacterial infections

Claim 3; Fig 3; 10pp; English.

This sequence is a Moraxella catarrhalis BASB019 protein of the
invention. The sequences can be used for diagnosis of disease, staging of
disease, or determining response of an infectious organism to drugs. The
polynucleotides may be used as a source for hybridisation probes, and for
screening of genetic mutations, serotype, organism or strain
identification, identification of mutation in BASB019 sequences, and as
components of arrays which are useful for diagnostic and prognostic
purposes. The polypeptides can be used to produce antibodies, and as a
target for the screening of antimicrobial drugs. The polypeptides can
also be used in vaccine formulations, and to identify agonists and
antagonists. The polypeptides, antibodies, agonists and antagonists
(which are bacteriostatic) are used for the treatment and prevention of
diseases including bacterial infection, otitis media in infants and

CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of Moraxella catarrhalis infections has risen dramatically,
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.

SO Sequence 172 AA;

Query Match 100.0%; Score 172; DB 21; Length 172;

Best Local Similarity 100.0%; Pred. No. 1.6e-154; Indels 0; Gaps 0;

Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLHIQIAAAALSVLTFTMGCAKSTQVMAVAPNAPGYGVYTVGAPLVNDERYK 60
DB 1 MMLHIQIAAAALSVLTFTMGCAKSTQVMAVAPNAPGYGVYTVGAPLVNDERYK 60
QY 61 ALASKLPFLVYFDFPSDEIKFQAAAILDEQAQFLTTNOTARVAVAGHTDERGSRREYNMSL 120
DB 61 ALASKLPFLVYFDFPSDEIKFQAAAILDEQAQFLTTNOTARVAVAGHTDERGSRREYNMSL 120
QY 121 GERRAVAVRNTLLGKINQASVEIISFGEERPIAFGTNEAMSQNRRAELSY 172
DB 121 GERRAVAVRNTLLGKINQASVEIISFGEERPIAFGTNEAMSQNRRAELSY 172

RESULT 2

AAV55090 standard; Protein: 172 AA.

AAV55090;

01-MAR-2000 (first entry)

M. catarrhalis BASB019 protein sequence #2.

BASB019 protein: diagnosis; infectious organism; auditive nerve damage;
genetic mutation screening; antibody production; vaccine; otitis media;
bacterial infection; pneumonia; sinusitis; nosocomial infection;
invasive disease; delayed speech learning; bacteria adhesion prevention;
upper respiratory tract infection; middle ear infection; therapy.

Moraxella catarrhalis.

WO9957277-A2.

11-NOV-1999.

03-MAY-1999; 99WO-EP03038.

06-MAY-1998; 98GB-0009683.

(SMIRK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-062148/05.

N-PSDB: AAZ40352.

Novel BASB019 polynucleotides and polypeptides from Moraxella

catarrhalis used to prepare vaccines against bacterial infections

Claim 3; Fig 3; 10pp; English.

This sequence is a Moraxella catarrhalis BASB019 protein of the

CC invention. The sequences can be used for diagnosis of disease, staging of
 CC disease, or determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and for
 CC screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies, and as a
 CC target for the screening of antimicrobial drugs. The polypeptides can
 CC also be used in vaccine formulations, and to identify agonists and
 CC antagonists. The polypeptides, antibodies, agonists and antagonists
 CC (which are bacteriostatic) are used for the treatment and prevention of
 CC diseases including bacterial infection, otitis media in infants and
 CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed speech
 CC learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarrhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarrhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used screen for new antibacterial compounds that may target these
 CC resistant bacteria.

CC Sequence 172 AA;

Query Match 62.2%; Score 107; DB 21; Length 172;
 Best Local Similarity 100.0%; Pred. No. 4.4e-93;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 66 LPSLVYDFDSDDEIKFOAAIIDEQAFITNOTARVLYAGHTDGRSGREYNMISLGERRA 125
 DB 66 LPSLVYDFDSDDEIKFOAAIIDEQAFITNOTARVLYAGHTDGRSGREYNMISLGERRA 125
 OY 126 VAVRNTLLGKGINQASVEITISFGEERPIAFGTNEBAMSQNRRAELSY 172
 DB 126 VAVRNTLLGKGINQASVEITISFGEERPIAFGTNEBAMSQNRRAELSY 172

RESULT 3

AAVS5091
 ID AAVS5091 standard; Protein: 172 AA.

AC AAVS5091;

DT 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein sequence #3.

XX BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
 KW genetic mutation screening; antibody production; vaccine; otitis media;
 KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
 KW invasive disease; delayed speech learning; bacteria adhesion prevention;
 KW upper respiratory tract infection; middle ear infection; therapy.

OS Moraxella catarrhalis.

PN WO9957277-A2.

PD 11-NOV-1999.

PF 03-MAY-1999; 99WO-EP03038.

PR 06-MAY-1998; 98GB-0009683.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PA Ruelle J;

XX WPI; 2000-062148/05.

DR N-PSDB; AA40353.
 XX
 PT Novel BASB019 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections
 PS Claim 3; Fig 3; 101pp; English.

CC This sequence is a Moraxella catarrhalis BASB019 protein of the
 CC invention. The sequences can be used for diagnosis of disease, staging of
 CC disease, or determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and for
 CC screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies, and as a
 CC target for the screening of antimicrobial drugs. The polypeptides can
 CC also be used in vaccine formulations, and to identify agonists and
 CC antagonists. The polypeptides, antibodies, agonists and antagonists
 CC (which are bacteriostatic) are used for the treatment and prevention of
 CC diseases including bacterial infection, otitis media in infants and
 CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed speech
 CC learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarrhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarrhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used screen for new antibacterial compounds that may target these
 CC resistant bacteria.

CC Sequence 172 AA;

Query Match 62.2%; Score 107; DB 21; Length 172;
 Best Local Similarity 100.0%; Pred. No. 4.4e-93;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 66 LPSLVYDFDSDDEIKFOAAIIDEQAFITNOTARVLYAGHTDGRSGREYNMISLGERRA 125
 DB 66 LPSLVYDFDSDDEIKFOAAIIDEQAFITNOTARVLYAGHTDGRSGREYNMISLGERRA 125
 OY 126 VAVRNTLLGKGINQASVEITISFGEERPIAFGTNEBAMSQNRRAELSY 172
 DB 126 VAVRNTLLGKGINQASVEITISFGEERPIAFGTNEBAMSQNRRAELSY 172

RESULT 4

AAVS5092
 ID AAVS5092 standard; Protein: 172 AA.

AC AAVS5092;

DT 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein sequence #4.

XX BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
 KW genetic mutation screening; antibody production; vaccine; otitis media;
 KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
 KW invasive disease; delayed speech learning; bacteria adhesion prevention;
 KW upper respiratory tract infection; middle ear infection; therapy.

OS Moraxella catarrhalis.

PN WO9957277-A2.

PD 11-NOV-1999.

PF 03-MAY-1999; 99WO-EP03038.

KW antibacterial; immune response; immunisation; vaccine.
 XX
 OS Synthetic.
 OS Moraxella catarrhalis.
 XX
 PN W0958684-A2.
 XX
 PD 18-NOV-1999.
 XX
 PF 07-MAY-1999; 99WO-EP03257.
 XX
 PR 13-MAY-1998; 98GB-0010285.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Rhonard J;
 XX
 DR WPI; 2000-062301/05.
 XX
 PT Novel peptides useful as vaccines for Moraxella infections such as
 PT otitis media, pneumonia, sinusitis etc.,
 XX
 PS Example 9; Page 69; 113pp; English.
 XX
 CC The present sequence represents a peptide specific for a BASB020
 CC protein, isolated from Moraxella catarrhalis, used in an example from
 CC the present invention. BASB020 polynucleotide fragments may be used as
 CC hybridisation probes for RNA, cDNA and genomic DNA to isolate full-length
 CC cDNAs and genomic clones encoding BASB020 and to isolate cDNA and genomic
 CC clones of other genes that have high sequence identity to BASB020 gene.
 CC The BASB020 polynucleotides and polypeptides are used as research
 CC reagents and materials for discovery of treatments of and diagnostics for
 CC human diseases. Probes comprising BASB020 nucleotide sequences can be
 CC constructed to conduct efficient screening of genetic mutations,
 CC serotype, taxonomic classification or identification. The polynucleotide
 CC sequences can be used in the discovery and development of antibacterial
 CC compounds. The polypeptides and polynucleotides are used to block the
 CC initial physical interaction between a gram negative and/or gram positive
 CC bacteria to the mammalian host. The polynucleotides encoding certain
 CC non-variable regions of bacterial cell surface protein are used in
 CC polynucleotide constructs which are useful for genetic immunisation
 CC experiments in animal models of infection with M. catarrhalis to
 CC identify protein epitopes able to provoke a prophylactic or therapeutic
 CC immune response. A therapeutic composition comprising an antibody
 CC directed against BASB020 can be used for treating humans with
 CC M. catarrhalis diseases such as sinusitis, otitis media and nosocomial
 CC infections.
 CC
 CC Sequence 16 AA:
 XX
 SQ
 Query Match 8.7%; Score 15; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 4.8e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 158 NEEAWSQNRRAELSY 172
 DB 2 NEEAWSQNRRAELSY 16
 RESULT 7
 AAY55094
 ID AAY55094 standard; Protein; 14 AA.
 XX
 AC AAY55094;
 XX
 DT 01-MAR-2000 (first entry)
 XX
 DE M. catarrhalis BASB019 protein sequence fragment.
 XX
 KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
 KW genetic mutation screening; antibody production; vaccine; otitis media;
 KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
 KW invasive disease; delayed speech learning; bacteria adhesion prevention;

KW upper respiratory tract infection; middle ear infection; therapy.
 XX
 OS Moraxella catarrhalis.
 XX
 PN W09957277-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 03-MAY-1999; 99WO-EP03038.
 XX
 PR 06-MAY-1998; 98GB-0009683.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Ruelle J;
 XX
 DR WPI; 2000-062148/05.
 XX
 PT Novel BASB019 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections
 XX
 PS Example 8; Page 70; 101pp; English.
 XX
 CC This sequence is a fragment of a Moraxella catarrhalis BASB019 protein of
 CC the invention. The sequences can be used for diagnosis of disease,
 CC staging of disease, or determining response of an infectious organism to
 CC drugs. The polynucleotides may be used as a source for hybridisation
 CC probes, and for screening of genetic mutations, serotype, organism or
 CC strain identification, identification of mutation in BASB013 sequences,
 CC and as components of arrays which are useful for diagnostic and
 CC prognostic purposes. The polypeptides can be used to produce antibodies,
 CC and as a target for the screening of antimicrobial drugs. The
 CC polypeptides can also be used in vaccine formulations, and to identify
 CC agonists and antagonists. The polypeptides, antibodies, agonists and
 CC antagonists (which are bacteriostatic) are used for the treatment and
 CC prevention of diseases including bacterial infection, otitis media in
 CC infants and children, pneumonia in the elderly, sinusitis, nosocomial
 CC infections and invasive diseases, chronic otitis media with hearing loss,
 CC fluid accumulation in the middle ear, auditive nerve damage, delayed
 CC speech learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarrhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarrhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used screen for new antibacterial compounds that may target these
 CC resistant bacteria.
 CC
 CC Sequence 14 AA:
 XX
 SQ
 Query Match 8.1%; Score 14; DB 21; Length 14;
 Best Local Similarity 100.0%; Pred. No. 3.7e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 46 YTGAPLVNDDEV 59
 DB 1 YTGAPLVNDDEV 14
 RESULT 8
 AAY57578
 ID AAY57578 standard; peptide; 14 AA.
 XX
 AC AAY57578;
 XX
 DT 02-MAR-2000 (first entry)
 XX
 DE Moraxella catarrhalis BASB020 specific peptide SEQ ID NO:14.
 XX
 KW Moraxella catarrhalis; BASB020; diagnosis; vaccine; infection;

PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR N-PSDB; AAS86383.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 52555; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 139 AA:

Query Match 4.7%; Score 8; DB 22; Length 139;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAAALS 15
Db 14 AAAAAAALS 21

RESULT 11
ID AAG18977 standard; Protein; 158 AA.
XX
AC AAG18977;
XX
DT 17-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 20601.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
OS
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130044.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139869.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141267.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.

PS Disclosure: SEQ ID NO 13455; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SO Sequence 306 AA:

Query Match 4.7%; Score 8; DB 22; Length 306;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAAATAAL 14

Db 195 TAAATAAL 202

RESULT 13

ABG11084

ID ABG11084 standard; Protein: 436 AA.

XX

AC ABG11084;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #11075.

XX

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WC200175067-A2.

XX

PD 11-OCT-2001.

XX

PE 30-MAR-2001; 2001MO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI: 2001-639362/73.

DR N-PSDB; AAS75271.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX

PS Claim 20; SEQ ID NO 41443; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (II) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SO Sequence 436 AA:

Query Match 4.7%; Score 8; DB 22; Length 436;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15

Db 13 AAAAAAALS 20

RESULT 14

AAV74388

ID AAV74388 standard; Protein: 452 AA.

XX

AC AAV74388;

XX

DT 21-MAR-2000 (first entry)

XX

DE *Neisseria meningitidis* ORF 081 protein sequence SEQ ID NO:252.

XX

KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

XX

OS *Neisseria meningitidis*.

XX

PN WC957280-A2.

XX

PD 11-NOV-1999.

XX

PE 30-APR-1999; 99MO-US09346.

XX

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098894.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX

PA (CHIR) CHIRON CORP.

PA (GENO-) INSTR GENOMIC RES.

XX

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,

PI Petersen J, Pizzo M, Rappelli R, Ratti G, Scalato E, Scarselli M,

PI Tettelin H, Venter JC;

XX

DR WPI: 2000-062150/05.

DR N-PSDB; AAZ53150.

XX

PT Novel *Neisseria* polypeptides predicted to be useful antigens for

PT vaccines and diagnostics.

XX

PS Claim 2; Page 267; 1453pp; English.

XX

CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAV74253 to AAV75941

CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides

CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent

CC PCR primers used in the exemplification of the present invention. The

CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX Sequence 452 AA;

Query Match 4.7%; Score 8; DB 21; Length 452;
Best Local Similarity 100.0%; Pred. No. 39;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
| | | | | | | |

Db 289 AAAAAALS 296

RESULT 15

ABBS578
ID ABB58578 standard; Protein: 545 AA.

XX ABB58578;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 2526.

KW Drosophila: developmental biology; cell signalling; insecticide;

KM pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB: ABL02681.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 2526; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 545 AA;

Query Match 4.7%; Score 8; DB 22; Length 545;

Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
| | | | | | | |

Db 369 AAAAAALS 376

RESULT 16

ABB67264
ID ABB67264 standard; Protein: 545 AA.

XX ABB67264;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 28584.

KW Drosophila: developmental biology; cell signalling; insecticide;

KM pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB: ABL11367.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 28584; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 545 AA;

Query Match 4.7%; Score 8; DB 22; Length 545;

Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
| | | | | | | |

Db 369 AAAAAALS 376

RESULT 17

AAU75084
ID AAU75084 standard; Protein: 606 AA.

XX AAU75084;

XX 23-APR-2002 (first entry)
 DT Ryegrass 4-coumarate Co-A-lyase 1 (LP4CL1) protein.
 DE
 XX
 KM Perennial ryegrass; lignin; 4 coumarate CoA-lyase; QTL;
 KM lignin biosynthesis; enzyme; cinamoyl-CoA reductase; CCR;
 KM cinamoyl alcohol dehydrogenase; CAD; molecular genetic marker;
 KM qualitative trait loci; tagging; QTL mapping; DNA fingerprinting;
 KM marker assisted selection; forage improvement; turf grass improvement;
 KM dry matter digestibility; herbage quality; palatability; regrowth;
 KM cold tolerance; drought tolerance; tiller survival; plant persistence.
 XX
 OS Lolium perenne.
 XX
 XX Location/Qualifiers
 FT Misc-difference 570..571
 FT /note="Encoded by GCCTGAAGA"
 XX
 PN MO200195702-A1.
 PD 20-DEC-2001.
 XX
 PF 14-JUN-2001; 2001WO-AU00699.
 XX
 PR 14-JUN-2000; 2000AU-0008154.
 XX
 PA (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.
 PA (UYAD-) UNIV ADELAIDE.
 PA (ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.
 PA (SACS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.
 PA (UYSC-) UNIV SOUTHERN CROSS.
 PA (DAIR-) DAIRY RES & DEV CORP.
 XX
 PI Spangenberg GC, Lidgett AJ, Heath RL, McInnes RL, Lynch DP;
 XX
 DR WPI: 2002-097993/13.
 DR N-PSDB: ABR13751.
 XX
 PT Novel nucleic acid encoding enzymes involved in lignin biosynthetic
 PT pathway from ryegrass or fescue species useful for modifying lignin
 XX biosynthesis in plants and as a molecular genetic marker -
 XX
 PS Claim 13; Fig 2; 148pp; English.
 CC
 CC This invention represents purified or isolated nucleic acid and protein
 CC sequences of enzymes involved in lignin biosynthesis. The enzymes
 CC of the invention are 4 coumarate CoA-lyase (4CL), cinamoyl-CoA
 CC reductase (CCR) and cinamoyl alcohol dehydrogenase (CAD) from a ryegrass
 CC (Lolium sp.) or fescue (Festuca sp.). The invention also comprises an
 CC isolated regulatory element from the nucleic acid sequences and a plant
 CC cell or seed transformed with the nucleic acid. An isolated regulatory
 CC element from these nucleotide molecules is useful for expressing an
 CC exogenous gene in plant cells. The nucleotide sequences of the invention
 CC and vectors containing these sequences are useful for modifying lignin
 CC biosynthesis in a plant and are useful as a molecular genetic marker for
 CC qualitative trait loci (QTL) tagging, QTL mapping, DNA fingerprinting
 CC and in marker assisted selection, in forage and turf grass improvement,
 CC e.g. tagging QTLs for dry matter digestibility, herbage quality,
 CC palatability, regrowth after cutting and grazing, cold tolerance,
 CC drought tolerance, tiller survival and plant persistence. The present
 CC sequence represents the perennial ryegrass 4-coumarate Co-A-lyase 1
 CC (LP4CL1) protein of the invention.
 CC
 SQ Sequence .606 AA;

Query Match 4.7%; Score 8; DB 23; Length 606;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 QIAAAAAA 13
 DB 13 QIAAAAAA 20

RESULT 18
 ID ABB70398 standard; Protein; 624 AA.
 XX ABB70398;
 AC ABB70398;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 37986.
 XX
 KM Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO20011042-A2.
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB: ABL14501.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions -
 XX
 PS Disclosure; SEQ ID NO 37986; 21pp + Sequence Listing; English.
 CC
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL10840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 624 AA;

Query Match 4.7%; Score 8; DB 22; Length 624;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 AAAAAAALS 15
 DB 158 AAAAAAALS 165

RESULT 19
 ID ABB66879 standard; Protein; 637 AA.
 XX ABB66879;
 AC ABB66879;
 XX

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 27429.
 XX
 KM Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical.
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR N-PSDB; ABL10982.
 DR WPI: 2001-656860/75.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 27429; 21bp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX Sequence 637 AA;
 SQ
 Query Match 4.7%; Score 8; DB 22; Length 637;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 AAAAALS 15
 |||||
 Db 461 AAAAALS 468
 RESULT 20
 ABB81482
 ID ABB81482 standard; Protein; 644 AA.
 XX
 XX ABB81482;
 AC
 XX 30-AUG-2002 (first entry)
 DT
 XX Rice acetolactate acid synthase protein SEQ ID NO:1.
 DE
 XX Rice; Oryza sativa var. Kinmaze; acetolactate acid synthase; enzyme;
 KM herbicide resistance; pyrimidinylcarboxy-based herbicide; plant.
 XX
 XX Oryza sativa.
 OS
 XX WO200244385-A1.
 PN
 XX 06-JUN-2002.
 PD
 XX 16-NOV-2001; 2001WO-JP10014.
 PF
 XX 29-NOV-2000; 2000JP-0362630.
 PR
 XX (TSUB) KUMIAT CHEM IND CO LTD.

PA (NAG-) NAT INST AGROBIOLOGICAL SCI.
 PI Shimizu T, Nakayama I, Nagayama K, Fukuda A, Tanaka Y, Kaku K;
 XX WPI: 2002-490301/52.
 DR N-PSDB; ABN89399.
 DR
 XX Gene encoding acetolactate acid synthase, useful in providing new breeds
 PT of plants with high resistance against pyrimidinylcarboxy-based
 PT herbicides -
 PT
 XX Claim 1; Page 82-86; 96pp; Japanese.
 PS
 XX The present sequence represents acetolactate acid synthase (I) isolated
 CC from Oryza sativa var. Kinmaze (rice). (I) has resistance against
 CC pyrimidinylcarboxy (PC)-based herbicides as well as acetolactate acid
 CC synthase activity (I) can be used for providing plants with high
 CC resistance against PC-based herbicides.
 CC
 XX Sequence 644 AA;
 SQ
 Query Match 4.7%; Score 8; DB 23; Length 644;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 AAAAALS 15
 |||||
 Db 6 AAAAALS 13
 RESULT 21
 AAU10024
 ID AAU10024 standard; Protein; 644 AA.
 XX
 XX AAU10024;
 AC
 XX 08-MAY-2002 (first entry)
 DT
 XX Rice acetohydroxyacid synthase (AHS) wild type protein sequence.
 DE
 XX Rice; AHS; acetohydroxyacid synthase; sulphonylurea herbicide;
 KM herbicide resistance; weed control; imidazolinone; EC.4.1.3.18.
 XX
 XX Oryza sativa.
 OS
 XX WO200185970-A2.
 PN
 XX 15-NOV-2001.
 PD
 XX 09-MAY-2001; 2001WO-US15072.
 PF
 XX 10-MAY-2000; 2000US-203434P.
 PR
 XX (LOU) UNIV LOUISIANA STATE & AGRIC & MECH COLL.
 PA
 XX Croughan TP;
 PI
 XX WPI: 2002-097559/13.
 DR N-PSDB; ABK14657.
 DR
 XX Novel gene encoding a functional acetohydroxyacid synthase gene which
 PT imparts resistance to at least one herbicide, used for producing
 PT herbicide resistant rice -
 PT
 XX Example 29; Page 117-119; 157pp; English.
 PS
 XX This invention relates to novel nucleic acid and protein sequences of
 CC mutant acetohydroxyacid synthase (AHS) enzymes that can be used to
 CC create herbicide resistance green plants. The encoded AHS exhibits
 CC resistance to at least one herbicide as compared to the wild-type, and
 CC has a serine-asparagine substitution at amino acid 627. The sequences of
 CC the invention are useful creating herbicide resistance plants, by
 CC planting these plants it is easier to control the growth of weeds in the

CC vicinity of a plant by applying a herbicide which normally inhibits AHAS.
CC The sequences can be used for producing rice plants having resistance to
CC at least one herbicide which normally inhibits AHAS in the wild-type
CC plant, such herbicides may be particularly imidazolinone or
CC sulphonylurea herbicides. The production of plant having resistance to
CC AHAS inhibiting herbicide allows the development of new herbicides which
CC target AHAS, reducing the risk of weeds becoming resistant. The present
CC sequence represents the wild type protein sequence of the rice AHAS
CC enzyme used to create the herbicide resistant plants of the invention.
XX
SQ Sequence 644 AA;

Query Match 4.7%; Score 8; DB 23; Length 644;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
DB 6 AAAAAALS 13

RESULT 22

AAU10025
ID AAU10025 standard; Protein; 644 AA.

AC AAU10025;

DT 08-MAY-2002 (first entry)

DE Rice herbicide resistant AHAS protein sequence.

KW Rice; AHAS; acetohydroxyacid synthase; sulphonylurea herbicide;
KW herbicide resistance; weed control; imidazolinone; EC.4.1.3.18;
XX mutant; muteln.

OS Oryza sativa.

PN WO200185970-A2.

PD 15-NOV-2001.

PF 09-MAY-2001; 2001WO-US15072.

PR 10-MAY-2000; 2000US-203434P.

PA (LOU) UNIV LOUISIANA STATE & AGRIC & MECH COLL.

PI Croughan TP;

DR WPI; 2002-097559/13.

DR N-PSDB; ABK14658.

PT Novel gene encoding a functional acetohydroxyacid synthase gene which
PT imparts resistance to at least one herbicide, used for producing
PT herbicide resistant rice.

XX Disclosure: Page 121-123; 157pp; English.

CC This invention relates to novel nucleic acid and protein sequences of
CC creatine acetohydroxyacid synthase (AHAS) enzymes that can be used to
CC create herbicide resistance green plants. The encoded AHAS exhibits
CC resistance to at least one herbicide as compared to the wild-type, and
CC has a serine-asparagine substitution at amino acid 627. The sequences of
CC the invention are useful creating herbicide resistance plants, by
CC planting these plants it is easier to control the growth of weeds in the
CC vicinity of a plant by applying a herbicide which normally inhibits AHAS.
CC The sequences can be used for producing rice plants having resistance to
CC at least one herbicide which normally inhibits AHAS in the wild-type
CC plant, such herbicides may be particularly imidazolinone or
CC sulphonylurea herbicides. The production of plant having resistance to
CC AHAS inhibiting herbicide allows the development of new herbicides which
CC target AHAS, reducing the risk of weeds becoming resistant. The present
CC sequence represents the sequence of the mutant protein sequence of the

CC rice AHAS enzyme used to create the herbicide resistant plants of the
CC invention.

SQ Sequence 644 AA;

Query Match 4.7%; Score 8; DB 23; Length 644;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
DB 6 AAAAAALS 13

RESULT 23

ABB59430
ID ABB59430 standard; Protein; 662 AA.

AC ABB59430;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 5082.

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR N-PSDB; ABL03533.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure: SEQ ID NO 5082; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB557737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 662 AA;

Query Match 4.7%; Score 8; DB 22; Length 662;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
DB 210 AAAAAALS 217

RESULT 24
ABB57802 standard; Protein; 1164 AA.
ID ABB57802;
XX ABB57802;
AC ABB57802;
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 198.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PE
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
XX N-PSDB; ABL01905.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 198; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 1164 AA;
SQ
Query Match 4.7%; Score 8; DB 22; Length 1164;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 AAAAAAALS 15
DB 940 AAAAAAALS 947
IIIIIIII
RESULT 25
ABB60943 standard; Protein; 1412 AA.
ID ABB60943;
XX ABB60943;
AC ABB60943;
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 9621.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX

XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PE
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
XX N-PSDB; ABL03046.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 9621; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175), and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 1412 AA;
SQ
Query Match 4.7%; Score 8; DB 22; Length 1412;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 AAAAAAALS 15
DB 978 AAAAAAALS 985
IIIIIIII
RESULT 26
ABB64682 standard; Protein; 2090 AA.
ID ABB64682;
XX ABB64682;
AC ABB64682;
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 20838.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PE
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA

XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX N-PSDB: ABL08785.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure: SEQ ID NO 20838; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (AB57737-AB572072).
XX (AB57737-AB572072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2090 AA:
SQ
Query Match 4.7%; Score 8; DB 22; Length 2090;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 AAAAAAALS 15
DB 1349 AAAAAAALS 1356
RESULT 27
AB563299
ID ABB63299 standard; Protein; 2703 AA.
XX
XX ABB63299;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 16689.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
XX
XX pharmaceutical.
KW
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PK
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI: 2001-656860/75.
DR
XX
XX N-PSDB: ABL07402.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure: SEQ ID NO 16689; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB57737-AB572072).
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
CC Sequence 2703 AA:
SQ
Query Match 4.7%; Score 8; DB 22; Length 2703;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 AAAAAAALS 15
DB 2647 AAAAAAALS 2654
RESULT 28
AAW04637
ID AAW04637 standard; Peptide; 9 AA.
XX
XX AAW04637;
AC
XX
XX 08-AUG-1997 (first entry)
DT
XX
XX Null peptide that complexes with A2 protein.
DE
XX
XX primer: PCR: polymerase chain reaction; specific inhibitor;
KW
XX
XX T-cell receptor; TCR: MHC; ligand; autoimmune disease; diabetes;
KW
XX
XX rheumatoid arthritis; Grave's disease; organ transplant rejection;
KW
XX
XX antigen; TCR-zeta; chimeric; CTL clone; HLA A2.1; HIV pol.
XX
XX
XX Synthetic.
OS
XX
XX WO9636881-A2.
PN
XX
XX 21-NOV-1996.
PD
XX
XX 16-MAY-1996; 96WO-GB01165.
PF
XX
XX 16-MAY-1995; 95GB-0009844.
PR
XX
XX (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PA
XX
XX Barouch DH, Jakobsen BK, Vessey SJR;
PI
XX
XX WPI: 1997-012213/01.
DR
XX
XX Specific inhibitors of interaction between T cell receptor and MHC
XX peptide ligand - identified by incubating receptor expressing cells
XX with ligand and test cpd., and measuring change in interaction to
XX detect cpds. potentially useful for blocking disease related T cells
XX
XX Example -: Page 3; 21pp; English.
PS
XX
XX Novel specific inhibitors of the interaction between a T-cell receptor
XX (TCR) and a MHC peptide ligand, are identified by: (a) incubating
XX responder cells that express TCR with a MHC ligand which stimulates the
XX cells, and a test cpd.; (b) monitoring a signal produced by the cell
XX when the TCR and MHC peptide ligand interact; and (c) comparing this
XX signal with a control signal from a similar system lacking the test
XX cpd.. The specific inhibitors are potentially useful for blocking
XX T-cells that cause autoimmune diseases (e.g. diabetes, rheumatoid
XX arthritis, Grave's disease etc.), organ transplant rejection or other T
XX cell mediated conditions. The interaction between a specific MHC peptide
XX ligand and a TCR was studied using purified MHC-single peptide complexes
XX (e.g. AAW04633-37) and an antigen specific TCR-zeta chimeric receptor
XX expressed on the surface of a basophil cell line. AAT43774-77 are primer

CC sequences used to amplify TCR genes by PCR from cDNA prepared from a
 CC CTL clone specific for HLA A2.1 restricted HIV pol peptide.
 XX
 SO Sequence 9 AA;

Query Match 4.1%; Score 7; DB 18; Length 9;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 DB 3 AAAAAA 9

RESULT 29
 AAW61561
 ID AAW61561 standard; peptide: 10 AA.

XX AAW61561;

AC 19-OCT-1998 (first entry)

DE Biotinylated peptide GYR(A)6L.

XX T-cell; malaria; immunogenic; anti-malarial; prophylactic immunity.

OS Synthetic.

PM WO9831382-A1.

PD 23-JUL-1998.

PF 21-JAN-1998; 98WO-US01527.

PR 21-JAN-1997; 97US-0033916.

PA (UYNV) UNIV NEW YORK STATE.

PI Moreno A, Nardin E;

DR WPI: 1998-413810/35.

PT New immunogenic compositions for malaria - comprise malaria derived
 peptide comprising universal T-cell epitope which elicits
 anti-malarial T-cell response

PS Disclosure: Page 11; 38pp; English.

CC The biotinylated peptides AAW61559-W61561 were used in a peptide binding
 assay to test the peptide interactions of a T-cell epitope derived from
 CC malaria can be used in an immunogenic composition. The T-cell epitope
 elicits an anti-malarial T-cell response in mammals of diverse genetic
 CC backgrounds. The composition can be used as a vaccine to confer
 CC prophylactic or therapeutic immunity against malaria. They may also be
 CC used to inhibit the propagation of a malarial organism in a susceptible
 CC animal.

SQ Sequence 10 AA;

Query Match 4.1%; Score 7; DB 19; Length 10;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 DB 4 AAAAAA 10

RESULT 30
 AAR84046
 ID AAR84046 standard; peptide: 13 AA.
 XX
 AC AAR84046;

XX 14-MAY-1996 (first entry)

DE Murine MHC class II binding peptide E5:10.

XX Murine; MHC class II; binding peptide; haptenated peptides;
 KW contact; sensitivity; desensitising; mammal; allergen; ivy;
 KW urushiol; poison; oak.

OS Synthetic.

Key	Location/Qualifiers
Modified-site	1
FT	/note="acylated"
FT	13
Modified-site	/note="amidated"

PN WO9526980-A2.

PD 12-OCT-1995.

PF 30-MAR-1995; 95WO-US04121.

PR 06-FEB-1995; 95US-0383645.

PR 01-APR-1994; 94US-0222206.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Geffer ML, Gelber C, Greenstein JL, Hackett CJ;

PI Wilson KJ;

DR WPI: 1995-358583/46.

PT Haptenated peptide(s) capable of binding to Class II MHC molecules -
 for treating contact dermatitis

PS Example: Fig 2; 85pp; English.

CC A peptide of 7-30 amino acids capable of binding to a murine MHC
 CC class II mol. (i.e. AAR84018-47) covalently linked to 1-3 hapten
 CC mols. can be used for treating contact sensitivity, or
 CC desensitising a mammal to a contact allergen (e.g. urushiol of
 CC poison ivy/oak). The peptide-hapten opds. disrupt the normal
 CC proliferation of hapten-specific T cells, or alter the T cell
 CC mediated delayed-type hypersensitivity response to the hapten,
 CC resulting in effective desensitisation to the hapten.

SQ Sequence 13 AA;

Query Match 4.1%; Score 7; DB 16; Length 13;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 IAAAAA 13
 DB 3 IAAAAA 9

RESULT 31
 AAW02569
 ID AAW02569 standard; peptide: 14 AA.

AC AAW02569;

DE 13-JAN-1997 (first entry)

DE AKAP79 A37-50 mutation.

XX Autoimmune disease; PACT59; PACT74; PACT36; PACT60; murine; T-cell; PKA;
 KW mouse; A-kinase anchoring protein 79; CAMP-dependent protein kinase;
 KW postsynaptic density; AKAP79; human forearm; transcriptional activator;
 KW interleukin 2; T-cell activation; calcineurin; T-cell clonal expansion;
 KW calcium/calmodulin dependent protein phosphatase; T-cell response;

KW autoimmune related disease; therapy; immune response.
 XX Synthetic.
 XX
 FT Key Location/Qualifiers
 FT MISC-difference 3
 FT MISC-difference 4 /note= "Arg39Ala"
 FT MISC-difference 5 /note= "Arg40Ala"
 FT MISC-difference 6 /note= "Lys41Ala"
 FT MISC-difference 9 /note= "Lys42Ala"
 FT MISC-difference 14 /note= "Lys45Ala"
 FT MISC-difference 14 /note= "Lys50Ala"
 XX
 PN W09616172-A2.
 PD 30-MAY-1996.
 XX
 PF 22-NOV-1995; 95MO-US16039.
 XX
 PR 17-JUL-1995; 95US-0503226.
 PR 23-NOV-1994; 94US-0344227.
 PR 15-MAR-1995; 95US-0404731.
 XX
 PA (ICOS-) ICOS CORP.
 PA (OREG-) STATE OF OREGON.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 PI Cognhan VM, Gallatin WM, Howard ML, Lockerbie RO;
 PI Scott JD;
 XX
 DR WPI; 1996-268608/27.
 XX
 PT New modulators of anchoring protein function - used to develop
 PT prods. for use in the treatment of auto-immune-related conditions.
 XX
 PS Example 11; Page 34; 74pp; English.
 XX
 CC AAM02565-W02569 represent mutants of residues 37-50 of the A-kinase
 CC anchoring protein 79 (AKAP79). AKAP79 (see AAM05264 for wild type
 CC residues 37-50) is responsible for anchoring protein kinase
 CC (PKA) to specific intracellular sites. AKAP79 is predominantly present
 CC in postsynaptic densities in the human forebrain. The pathways that
 CC involve AKAP79 are important in many cell types and have been implicated
 CC in many cell functions, including the transcriptional activation of the
 CC interleukin 2 gene that is important in T-cell activation. AKAP also
 CC binds to calcineurin (see AAM02536), which is a calcium/calmodulin
 CC dependent protein phosphatase associated with T-cell activation. By
 CC binding both PKA and calcineurin, AKAP79 co-localises a kinase and a
 CC phosphatase which may regulate flux through a specific signalling
 CC pathway. The AKAP79 binding sequences can be used to develop products
 CC for use in the treatment of autoimmune related conditions. The AKAP79
 CC binding proteins can be used in methods for stimulating an immune
 CC response, and for stimulating activated T-cells for selected clonal
 CC expansion. The proteins can also be used in a method for enhancing
 CC T-cell responses to experimental stimuli for evaluation of early events
 CC in T-cell biology and activation of the immune response.
 CC
 SQ Sequence 14 AA;
 XX
 QY Query Match 4.1%; Score 7; DB 17; Length 14;
 DB Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 AAAAAAL 14
 DB 5 AAAAAAL 11

RESULT 32
 AAB14909
 ID AAB14909 standard; Peptide; 14 AA.
 XX
 AC AAB14909;
 XX
 DT 08-JAN-2001 (first entry)
 XX
 DE Mutant peptide AKAP 79 A37-50.
 XX
 KW Human; calcineurin-binding peptide; A-kinase anchor protein; AKAP;
 KW AKAP 79; immunostimulant; interleukin 2 expression modulation;
 KW graft rejection; transplantation; T cell-mediated disorder; mutant.
 XX
 OS Homo sapiens...
 OS Synthetic.
 XX
 PN US6107104-A.
 PD 22-AUG-2000.
 XX
 PF 27-SEP-1996; 96US-0721458.
 XX
 PR 23-NOV-1994; 94US-0344227.
 PR 15-MAR-1995; 95US-0404731.
 PR 17-JUL-1995; 95US-0503226.
 XX
 PA (ICOS-) ICOS CORP.
 XX
 PI Lockerbie RO, Gallatin WM, Lai Y, Howard ML;
 XX
 DR WPI; 2000-578541/54.
 XX
 PT Novel calcineurin deletion mutant having calcineurin polypeptide
 PT sequence and binding A-kinase anchor proteins, for treating graft
 PT rejection following organ transplantation and T cell-mediated disorders
 XX
 PS Example 11; Column 21; 53pp; English.
 XX
 CC The present sequence is a mutant peptide derived from A-kinase anchor
 CC protein 79 (AKAP 79). It is expressed as a poly-histidine tag fusion
 CC protein and can thus be purified to homogeneity by nickel affinity
 CC chromatography. AKAP 79 binds both CAMP-dependent protein kinase (PKA)
 CC and calcineurin and so co-localises a kinase and a phosphatase that
 CC may regulate flux through a specific signalling pathway. Calcineurin is a
 CC Ca2+/calmodulin-dependent protein phosphatase which is involved in many
 CC intracellular signalling pathways. It participates in regulation of IL-2
 CC expression following T cell stimulation in T cells. Calcineurin-binding
 CC peptides derived from AKAP 79 may be used to inhibit calcineurin activity
 CC in a cell. The peptides are useful for treating graft rejection following
 CC organ transplantation and for treating T cell-mediated disorders.
 CC Calcineurin deletion mutants which bind AKAP 79 are useful for defining
 CC an AKAP 79 binding site, for stimulating the immune response, stimulating
 CC activated T cells for selected clonal expansion, or for enhancing T cell
 CC responses to experimental stimuli for evaluation of early events in
 CC T cell biology and activation of the immune response.
 CC
 SQ Sequence 14 AA;
 XX
 QY Query Match 4.1%; Score 7; DB 21; Length 14;
 DB Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 AAAAAAL 14
 DB 5 AAAAAAL 11
 RESULT 33
 AAR84047
 ID AAR84047 standard; peptide; 16 AA.
 XX

```

AC AAR84047;
XX
XX 14-MAY-1996 (first entry)
DT
XX
XX Murine MHC class II binding peptide E5:10:PDC.
DE
XX
XX Murine; MHC class II; binding peptide; haptenated peptides;
KM contact; sensitivity; desensitising; mammal; allergen; ivy;
KW urushiol; poison; oak.
XX
XX Synthetic.
XX
XX Key location/Qualifiers
XX Modified-site 1 /note= "acylated"
XX Modified-site 16 /note= "amdated"
XX
XX WO9526980-A2.
XX
XX 12-OCT-1995.
XX
XX 30-MAR-1995; 95WO-US04121.
XX
XX 06-FEB-1995; 95US-0383645.
XX 01-APR-1994; 94US-0222206.
XX
XX (IMMUNO-) IMMUNOLOGIC PHARM CORP.
XX
XX Gefter ML, Gelber C, Greenstein JL, Hackett CJ;
XX Wilson KJ;
XX
XX WPI: 1995-358583/46.
XX
XX Haptenated peptide(s) capable of binding to Class II MHC molecules -
XX for treating contact dermatitis
XX
XX Example; Fig 2; 85pp; English.
XX
XX A peptide of 7-30 amino acids capable of binding to a murine MHC
XX class II mol. (i.e. AAR84018-47) covalently linked to 1-3 hapten
XX mols. can be used for treating contact sensitivity, or
XX desensitising a mammal to a contact allergen (e.g. urushiol of
XX poison ivy/oak). The peptide-hapten cpds. disrupt the normal
XX proliferation of hapten-specific T cells, or alter the T cell
XX mediated delayed-type hypersensitivity response to the hapten,
XX resulting in effective desensitisation to the hapten.
XX
XX Sequence 16 AA:
SQ
XX
XX Query Match 4.1%; Score 7; DB 16; Length 16;
XX Best Local Similarity 100.0%; Pred. No. 17;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 IAAAAA 13
XX |||||
DB 3 IAAAAA 9
XX

RESULT 34
ABB27942
ID ABB27942 standard; Peptide; 21 AA.
XX
XX ABB27942;
XX
XX 01-FEB-2002 (first entry)
DT
XX
XX Human peptide #593 encoded by breast cell single exon nucleic acid probe.
DE
XX
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS

```

```

XX
XX WO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00662.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 10910; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a peptide encoded by a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 21 AA:
SQ
XX
XX Query Match 4.1%; Score 7; DB 22; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 22;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 130 NYLGGK 136
XX |||||
DB 4 NYLGGK 10
XX

RESULT 35
ABB37770
ID ABB37770 standard; Peptide; 21 AA.
XX
XX ABB37770;
XX
XX 04-FEB-2002 (first entry)
DT
XX
XX Peptide #5276 encoded by human foetal liver single exon probe.
DE
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
KW Homo sapiens.
XX
XX WO200157271-A2.
OS

```

XX 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00669.
 XX
 PE
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 27; SEQ ID NO 30405; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human fetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 SO Sequence 21 AA;
 SQ
 Query Match 4.1%; Score 7; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 NYLGGK 136
 Db 4 NYLGGK 10
 XX
 RESULT 36
 AAM53912
 ID AAM53912 standard; Protein; 21 AA.
 XX
 AC AAM53912;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26017.
 XX
 KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PE 30-JAN-2001; 2001WO-US00667.
 XX
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483446/52.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 PS Example 4; SEQ ID NO: 26017; 650pp + Sequence listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 CC
 SO Sequence 21 AA;
 SQ
 Query Match 4.1%; Score 7; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 NYLGGK 136
 Db 4 NYLGGK 10
 XX
 RESULT 37
 AAM70864
 ID AAM70864 standard; Protein; 21 AA.
 XX
 AC AAM70864;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31170.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PE 30-JAN-2001; 2001WO-US00668.
 XX
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 31170; 658pp + Sequence listing; English.
 CC The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.

XX Sequence 21 AA;

SO Query Match 4.1%; Score 7; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 130 NYLKGK 136
Db 4 NYLKGK 10

RESULT 38
AAM18708 standard; Protein; 21 AA.

XX AAM18708;
XX 12-OCT-2001 (first entry)
XX Peptide #5142 encoded by probe for measuring cervical gene expression.
XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -

XX Claim 27; SEQ ID NO 23534; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
XX (SENP: see A4110068-A128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human Hela cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 21 AA;

Query Match 4.1%; Score 7; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 130 NYLKGK 136

Db 4 NYLKGK 10

RESULT 39

AAM26577 standard; Protein; 21 AA.

XX AAM26577;

XX 17-OCT-2001 (first entry)

XX Peptide #614 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID NO 26846; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP:
XX see A4131315-A157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.

XX Sequence 21 AA;

Query Match 4.1%; Score 7; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 130 NYLKGK 136

Db 4 NYLKGK 10

RESULT 40

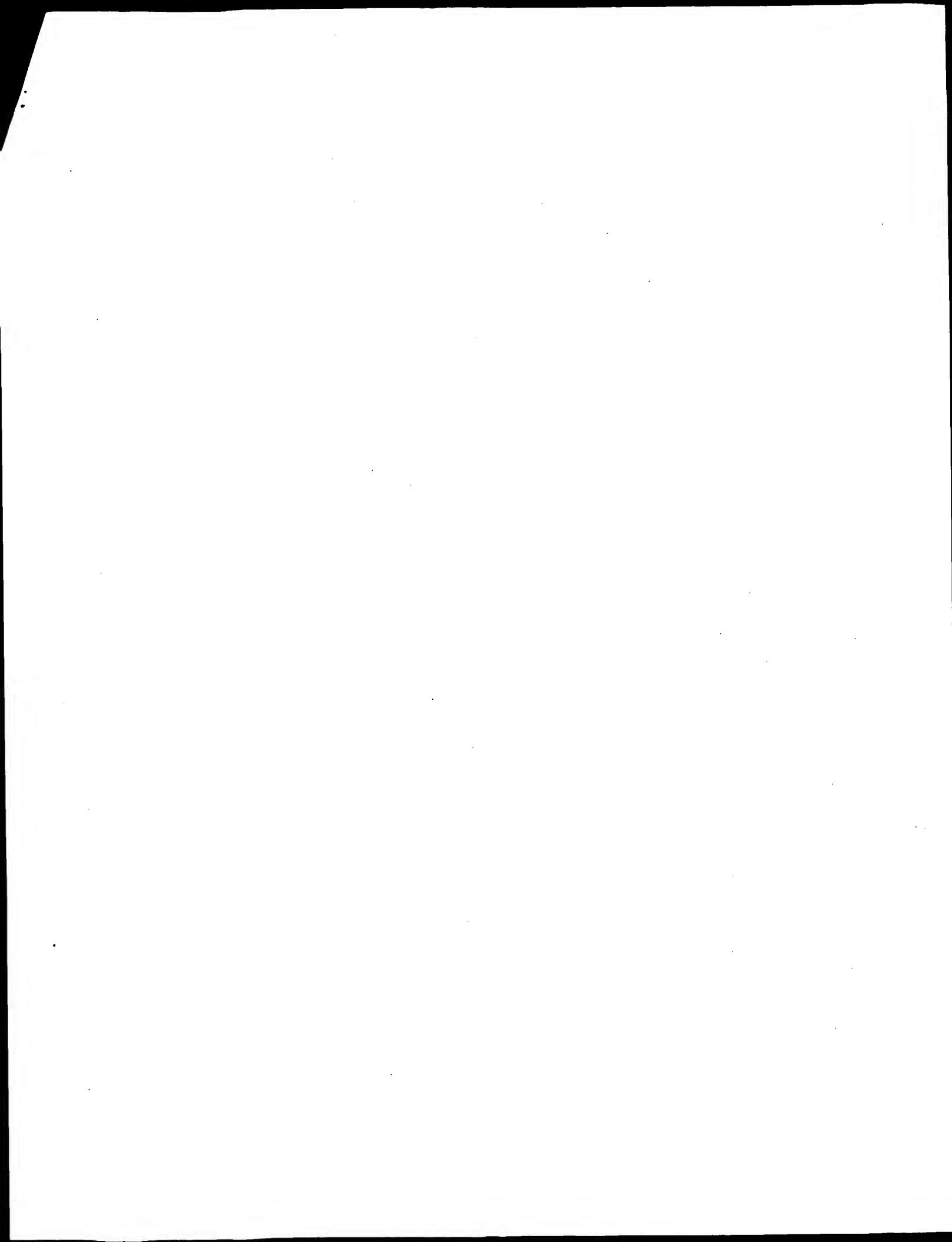
ABG40662 standard; Peptide; 21 AA.

XX ABG40662;

XX 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 30327.
 XX
 KW Human: single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 PS Claim 27; SEQ ID No 30327; 634pp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression to a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension

CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 21 AA;
 Query Match 4.18; Score 7; DB 23; Length 21;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 130 NYYLGGK 136
 4 NYYLGGK 10
 Search completed: July 6, 2003, 14:20:17
 Job time : 39 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:20:20 ; Search time 21 Seconds
(without alignments)
787,387 Million cell updates/sec

Title: US-09-674-779b-2

Perfect score: 172
Sequence: 1 MMLHTQIAAAALSVLTGM.....IAFGTEEAWSQNRRAELSY 172

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	7.6	176	2	A60337
2	10	5.8	186	2	E82625
3	9	5.2	2176	2	T13806
4	8	4.7	143	2	H72296
5	8	4.7	280	2	B95296
6	8	4.7	381	2	UC7650
7	8	4.7	452	2	H81777
8	8	4.7	491	2	I40455
9	8	4.7	526	2	G82981
10	8	4.7	547	2	T50939
11	8	4.7	829	2	A34692
12	8	4.7	2715	2	T13045
13	7	4.1	37	1	FDFL3W
14	7	4.1	40	1	FDFL3W
15	7	4.1	40	1	FDFL3W
16	7	4.1	82	2	PDFLAW
17	7	4.1	82	2	S02326
18	7	4.1	82	2	JS0706
19	7	4.1	82	2	A05161
20	7	4.1	82	2	JS0705
21	7	4.1	87	2	IS1125
22	7	4.1	150	2	A55846
23	7	4.1	165	2	F86299
24	7	4.1	167	2	B70533
25	7	4.1	167	2	S58217
26	7	4.1	168	2	G83525
27	7	4.1	168	2	I40346
28	7	4.1	168	2	AF3294
29	7	4.1	170	2	AE2405
30	7	4.1	173	1	LPECPG

30	7	4.1	173	2	A85577	peptidoglycan-asso
31	7	4.1	173	2	H90725	peptidoglycan-asso
32	7	4.1	174	2	A10592	peptidoglycan-asso
33	7	4.1	174	2	F72550	hypothetical prote
34	7	4.1	176	2	T52661	cysteine synthase
35	7	4.1	177	2	AE3013	omp6 protein [imp
36	7	4.1	177	2	B98271	omp6 protein [imp
37	7	4.1	179	2	D75351	probable acetyltra
38	7	4.1	185	2	A62752	hypothetical prote
39	7	4.1	188	2	H71495	probable peptidogl
40	7	4.1	192	2	E72036	peptidoglycan-asso
41	7	4.1	192	2	C65588	peptidoglycan-asso
42	7	4.1	195	2	AG2727	conserved hypotet
43	7	4.1	200	2	G86214	protein T6D22.4 [1
44	7	4.1	202	2	G81653	peptidoglycan asso
45	7	4.1	203	2	I49153	cardiotrophin-1 -
46	7	4.1	213	2	A84250	NADH oxidase [limp
47	7	4.1	214	1	JC4808	ribosomal protein
48	7	4.1	214	2	E75613	hypothetical prote
49	7	4.1	233	2	D95860	probable transcrip
50	7	4.1	240	2	E72629	hypothetical prote
51	7	4.1	244	2	A98330	sporulation transc
52	7	4.1	244	2	AD2953	transcription regu
53	7	4.1	251	2	B07675	probable cobM - My
54	7	4.1	260	2	C70675	hypothetical prote
55	7	4.1	260	2	E87310	pyroline-5-carbox
56	7	4.1	261	2	S63604	homeobox protein G
57	7	4.1	267	2	B97509	hypothetical prote
58	7	4.1	267	2	S38367	tetrahydromethanop
59	7	4.1	267	2	G69021	phosphate-binding
60	7	4.1	271	2	H69097	pyroline-5-carbox
61	7	4.1	272	2	AH2847	delta 1-pyrraline-
62	7	4.1	274	2	G97624	hypothetical prote
63	7	4.1	279	2	F84320	probable ATP-bindi
64	7	4.1	288	2	C81422	photosystem II oxy
65	7	4.1	291	2	S05508	ribokinase rbsk [1
66	7	4.1	294	2	H84115	hypothetical prote
67	7	4.1	296	2	T30575	metalio-beta-lacta
68	7	4.1	302	2	A12675	ribokinase [import
69	7	4.1	303	2	B87367	hypothetical prote
70	7	4.1	303	2	F84401	genomic screen hom
71	7	4.1	305	2	I57039	hydroxyacylgutath
72	7	4.1	306	2	AB3431	probable PE protei
73	7	4.1	308	2	D70875	sloppy paired prot
74	7	4.1	322	2	S23053	gas-vesicle operon
75	7	4.1	322	2	S28125	alpha-antigen prec
76	7	4.1	325	2	A37185	hypothetical prote
77	7	4.1	326	2	D72689	probable metallobe
78	7	4.1	326	2	H97457	Graves disease mit
79	7	4.1	330	2	S26596	probable serpin [1
80	7	4.1	331	2	E84770	homeotic protein E
81	7	4.1	333	2	A39065	lipopolysaccharide
82	7	4.1	335	2	S70671	hypothetical prote
83	7	4.1	335	2	B87590	probable signal pe
84	7	4.1	336	2	T34783	hypothetical prote
85	7	4.1	339	2	T26328	acidic ribosomal p
86	7	4.1	343	2	E64363	hypothetical prote
87	7	4.1	344	2	F70922	lichenase (EC 3.
88	7	4.1	349	2	A44507	REIC protein - huc
89	7	4.1	350	2	JC7188	hydrogenase (EC 1.
90	7	4.1	353	2	B70358	helix-loop-helix t
91	7	4.1	367	2	JC6087	nifs protein homol
92	7	4.1	370	2	E69754	hypothetical prote
93	7	4.1	374	2	B64475	probable aspartate
94	7	4.1	388	2	C70605	protein RFL - rice
95	7	4.1	389	2	T03411	conserved oxidoredu
96	7	4.1	394	2	E75439	hypothetical prote
97	7	4.1	394	2	H70605	homeotic protein H
98	7	4.1	401	2	T01001	hypothetical prote
99	7	4.1	403	2	A53662	
100	7	4.1	404	2	T24255	

ALIGNMENTS

RESULT 1

A60337

outer membrane protein pPIA, peptidoglycan-associated, precursor - Legionella pneumophila

N:Alternate names: lipoprotein antigen

C:Species: Legionella pneumophila

C>Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1993

C:Accession: A60337; S16531

R:Ludwig, B.; Schmid, A.; Marre, R.; Hacker, J.

R:Enleberg, N.C.; Howe, D.C.; Rogers, J.E.; Arroyo, J.; Eisenstein, B.I.

Mol. Microbiol. 5, 2021-2029, 1991

A:Title: Cloning, genetic analysis, and nucleotide sequence of a determinant coding for

A:Reference number: A60337; MUID:91310296; PMID:1855972

A:Accession: A60337

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <LUD>

A:Title: Characterization of a Legionella pneumophila gene encoding a lipoprotein antigen

A:Reference number: S16531; MUID:92114778; PMID:1766377

A:Accession: S16531

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <ENG>

A:Cross-references: EMBL:X60543; NID:g44128; PIDN:CAAA3033.1; PID:g44129

C:Superfamily: outer membrane protein A

C:Keywords: membrane protein; surface antigen

Query Match

Best Local Similarity 7.6%; Score 13; DB 2; Length 176;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 AGHDEGSRREYN 117

DB 105 AGHDEGSRREYN 117

RESULT 2

outer membrane protein p6 precursor XPI896 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: E82625

R:Anonimus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365177; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: E82625

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-186 <STM>

A:Cross-references: GB:AE004009; GB:AE003849; NID:9106980; PIDN:AA84702.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A

Briano, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H

as Neto, E.; Docena, C.; El-Dorri, H.; Facinanci, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; From

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

C:Contents: annotation

A:Gene: XPI896

Query Match 5.8%; Score 10; DB 2; Length 186;
 Best Local Similarity 100.0%; Pred. No. 0.065;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 TDEGSRREYN 117

DB 119 TDEGSRREYN 128

RESULT 3

toucan gene protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13806

R:Grammont, M.; Dastugue, B.; Couderc, J.L.

Development 124, 4917-4926, 1997

A:Title: The Drosophila toucan (tcc) gene is required in germline cells for somatic c

A:Reference number: Z17769; MUID:98090047; PMID:9362455

A:Accession: T13806

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2176 <GRA>

A:Cross-references: EMBL:Y14157; NID:g2760521; PIDN:CAA4574.1; PID:g2760522

C:Superfamily: FlyBase:FBgn0015600

A:Gene: tcc

A:Cross-references: FlyBase:FBgn0015600

Query Match

Best Local Similarity 5.2%; Score 9; DB 2; Length 2176;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 16

DB 973 AAAAALSV 981

RESULT 4

sugar-phosphate isomerase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: H72296

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.

M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: H72296

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-143 <ARN>

A:Cross-references: GB:AE001768; GB:AE00512; NID:g4981619; PIDN:AAD36157.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TMI080

C:Superfamily: galactoside O-acetyltransferase

Query Match

Best Local Similarity 4.7%; Score 8; DB 2; Length 143;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 NYLGGKI 137

DB 20 NYLGGKI 27

RESULT 5

conserved hypothetical protein SMA0520 [imported] - Sinorhizobium meliloti (strain 10

C:Species: Sinorhizobium meliloti

C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: B95296
 R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
 ..; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
 A:Reference number: A95262; MUID:21396509; PMID:11481432
 A:Accession: B95296
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-280 <KUR>
 A:Cross-references: GB:AE006469; PIDN:AAK64932.1; PID:g14523354; GSPDB:GN00165
 R:Galibert, F.; Pihan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 P.; Chalm, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelure,
 P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Sma0520
 A:Genome: plasmid

Query Match 4.7%; Score 8; DB 2; Length 280;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 DB 121 AAAAAAALS 128

RESULT 6

JC7650
 pectin lyase (EC 4.2.2.10) - *Aspergillus oryzae*
 C:Species: *Aspergillus oryzae*
 C:Date: 30-Jun-2001 #sequence, revision 30-Jun-2001 #text_change 02-Nov-2001
 C:Accession: JC7650; PC7125
 R:Kikamoto, N.; Yoshino-Yasuda, S.; Ohmiya, K.; Tsukagoshi, N.
 Biosci. Biotechnol. Biochem. 65, 209-212, 2001
 A:Title: Sequence analysis and overexpression of a pectin lyase gene (pell) from *Asperg*
 A:Reference number: JC7650; MUID:21119720; PMID:11272833
 A:Accession: JC7650
 A:Molecule type: DNA
 A:Residues: 1-381 <KIT>
 A:Experimental source: strain KBN616
 A:Accession: PC7125
 A:Molecule type: protein
 A:Residues: 21-35 <KIT>
 C:Comment: This enzyme, a member of subclass of pectolytic enzymes, is the most important
 ta-elimination, to form galacturonic oligomers with 4,5-unsaturated residues at the non-
 C:Genetics:
 A:Gene: pell
 A:Introns: 159/22
 C:Keywords: carbon-oxygen lyase

Query Match 4.7%; Score 8; DB 2; Length 381;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAAAAAAL 14
 DB 7 TAAAAAAL 14

RESULT 7

UDP-N-acetylmutamoylalanyl-D-glutamyl-2,6-diamino-pimelate-D-alanyl-D-alanine ligase (EC
 C:Species: *Neisseria meningitidis*
 C:Date: 05-May-2000 #sequence, revision 05-May-2000 #text_change 03-Jun-2002
 C:Accession: H81777

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491
 A:Reference number: A81775; MUID:20222556; PMID:10761919
 A:Accession: H81777
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-452 <PAR>
 A:Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85286.1; PID:g738
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: muref; NMA2068
 C:Keywords: ligase

Query Match 4.7%; Score 8; DB 2; Length 452;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 DB 289 AAAAAAALS 296

RESULT 8

140455
 penicillin binding protein pbp - *Bacillus subtilis*
 C:Species: *Bacillus subtilis*
 C:Date: 12-Aug-1996 #sequence, revision 12-Aug-1996 #text_change 21-Jul-2000
 C:Accession: I40455; D69673; S49132
 R:Tognoni, A.; Franchi, E.; Magistrelli, C.; Colombo, E.; Cosmina, P.; Grandi, G.
 Microbiology 141, 645-648, 1995
 A:Title: A putative new peptide synthase operon in *Bacillus subtilis*: partial charact
 A:Reference number: I40454; MUID:95227362; PMID:7711903
 A:Accession: I40455
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-491 <RES>
 A:Cross-references: EMBL:Z34883; NID:g1805667; PIDN:CAB4366.1; PID:g509467
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C:Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A:Ethlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Fowler, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funo, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogihara, K.; Rocha, B.; Roche, B.; Rose, M.; Sadale, V.; Sato, T.; Scari
 Rieger, M.; Rivolta, C.; Rocha, B.; Roche, B.; Rose, M.; Sadale, V.; Sato, T.; Scari
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: D69673
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-491 <KUN>
 A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13718.1; PID:g26342
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: pbp
 C:Superfamily: D-alanyl-D-alanine carboxypeptidase

Query Match 4.7%; Score 8; DB 2; Length 491;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAAISVL 17
 DB 88 AAAAISVL 95

RESULT 9

682981
Probable binding protein component of ABC dipeptide transporter PA5317 [imported] - Pseu
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G82981
R:Stover, C.K.; Pham, X.O.; Errin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,
; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G82981
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-526 <STO>
A:Cross-references: GB:AE004944; GB:AE004091; NID:g9951628; PIDN:AA608702.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA5317
C:Superfamily: dipeptide transport protein

Query Match 4.7%; Score 8; DB 2; Length 526;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 AAAAALSV 16
|||||
Db 19 AAAAALSV 26

RESULT 10

750939
DITE protein [imported] - Pseudomonas abietaniphila
C:Species: Pseudomonas abietaniphila
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50939
R:Martin, V.J.; Mohu, W.W.
J: Bacteriol. 181, 2675-2682, 1999
A:Title: A novel aromatic-ring-hydroxylating dioxygenase from the diterpenoid-degrading
A:Reference number: Z25281; MUID:99235742; PMID:10217753
A:Accession: T50939
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-547 <MAR>
A:Cross-references: EMBL:AF119621; PIDN:AAD21067.1
A:Experimental source: strain BKME-9; ATCC700689
C:Genetics:
A:Gene: dITE

Query Match 4.7%; Score 8; DB 2; Length 547;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAALSVL 17
|||||
Db 107 AAAALSVL 114

RESULT 11

A34692
ecdysone-induced protein E74A - fruit fly (Drosophila melanogaster)
N:Alternate names: ets-related protein E74A
C:Species: Drosophila melanogaster
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 24-Sep-1998
C:Accession: A34692
R:Burtis, K.C.; Thummel, C.S.; Jones, C.W.; Karim, F.D.; Hogness, D.S.
Cell 61, 85-99, 1990
A:Title: The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene
A:Reference number: A90912; MUID:90199900; PMID:2107982
A:Accession: A34692
A:Molecule type: mRNA

A:Residues: 1-829 <BUR>
A:Cross-references: GB:M37082; NID:g157307; PID:g157308
C:Genetics:
A:Gene: E74
A:Cross-references: FlyBase:FBgn0000567
C:Superfamily: ets DNA-binding domain homology
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation
F:735-815/Domains: ets DNA-binding domain homology <ETS>

Query Match 4.7%; Score 8; DB 2; Length 829;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAAAAAL 14
|||||
Db 195 TAAAAAL 202

RESULT 12

T13049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13049
R:Freeman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z17592
A:Accession: T13049
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <TR>
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: FlyBase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the Wg pathway
C:Keywords: DNA binding

Query Match 4.7%; Score 8; DB 2; Length 2715;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALS 15
|||||
Db 2659 AAAAALS 2666

RESULT 13

FDPL3W
antifreeze protein 3 - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 23-Aug-1996
C:Accession: A03192
R:Devries, A.L.; Lin, Y.
Biochim. Biophys. Acta 495, 388-392, 1977
A:Title: Structure of a peptide antifreeze and mechanism of adsorption to ice.
A:Reference number: A03192; MUID:78060969; PMID:288591
A:Accession: A03192
A:Molecule type: protein
A:Residues: 1-37 <DEV>
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match 4.1%; Score 7; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
|||||
Db 6 AAAAAL 12

RESULT 14

FDRI8G
antifreeze protein GS-8 - grubby sculpin
A:Title: Differential amplification of antifreeze protein genes in the Pleuronectinae
A:Reference number: S02326; MUID:88259236; PMID:3133486
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 24-Nov-1999
C:Accession: S07046
R:Chakraborty, A.; Hew, C.L.; Shears, M.; Fletcher, G.
Can. J. Zool. 66, 403-408, 1988
A:Title: Primary structures of the alanine-rich antifreeze polypeptides from grubby sculpin
A:Reference number: S06417
A:Accession: S07046
A:Molecule type: protein
A:Residues: 1-40 <CH>
C:Superfamily: antifreeze protein
C:Keywords: antifreeze; blocked amino end
F:1/Modified site: blocked amino end (Met) #status experimental

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 14 AAAAAA 20

RESULT 15

FDPLAW

antifreeze protein A precursor - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 22-Jun-1999
C:Accession: JS0704; A03194
R:Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:92209995; PMID:1555765
A:Accession: JS0704
A:Molecule type: DNA
A:Residues: 1-82 <DA>
A:Cross-references: GB:M62412; GB:M62416; NID:g213592; PIDN:AAA49471.1; PID:g213593
R:Davies, P.L.; Roach, A.H.; Hew, C.L.
Proc. Natl. Acad. Sci. U.S.A. 79, 335-339, 1982
A:Title: DNA sequence coding for an antifreeze protein precursor from winter flounder.
A:Reference number: A03194; MUID:82197490; PMID:6952188
A:Accession: A03194
A:Molecule type: mRNA
A:Residues: 1-82 <DA>
A:Experimental source: clones 4-2b and 2A-7c
A>Note: the authors translated the codon AGC for residue 24 as Arg
C:Genetics:
A:Introns: 19/2
C:Superfamily: antifreeze protein
C:Keywords: antifreeze; plasma; tandem repeat
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-44/Domain: propeptide #status predicted <PRO>
F:45-82/Product: antifreeze protein A #status predicted <MAT>

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 50 AAAAAA 56

RESULT 16

S02326

antifreeze protein A - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 24-Oct-2000
C:Accession: S02326; JH0627
R:Scott, G.K.; Davies, P.L.; Kao, M.H.; Fletcher, G.L.

J. Mol. Evol. 27, 29-35, 1988

A:Title: Differential amplification of antifreeze protein genes in the Pleuronectinae
A:Reference number: S02326; MUID:88259236; PMID:3133486
A:Accession: S02326
A:Molecule type: DNA
A:Residues: 1-82 <SC>
A:Cross-references: EMBL:X07506; NID:964211; PIDN:CAA30389.1; PID:964212
R:Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:92209995; PMID:1555765
A:Accession: JH0627
A:Molecule type: DNA
A:Residues: 1-82 <DA>
A:Cross-references: GB:M62415
C:Genetics:
A:Introns: 19/2
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 50 AAAAAA 56

RESULT 17

JS0706

antifreeze protein (clone 4-2c) - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 24-Oct-2000
C:Accession: JS0706
R:Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:92209995; PMID:1555765
A:Accession: JS0706
A:Molecule type: DNA
A:Residues: 1-82 <DA>
A:Cross-references: GB:M62417; NID:g213594; PIDN:AAA49472.1; PID:g213595
A>Note: the authors translated the codon AGC for residue 24 as Arg
C:Genetics:
A:Introns: 19/2
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 50 AAAAAA 56

RESULT 18

A05161

antifreeze protein B precursor - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 24-Oct-2000
C:Accession: A05161
R:Davies, P.L.; Hough, C.; Scott, G.K.; Ng, N.; White, B.N.; Hew, C.L.
J. Biol. Chem. 259, 9241-9247, 1984
A:Reference number: A05161; MUID:84264559; PMID:6086629
A:Accession: A05161
A:Molecule type: DNA
A:Residues: 1-82 <DA>
A:Cross-references: GB:L00138; GB:J00929; NID:g343126; PIDN:AAB59964.1; PID:g457351
C:Genetics:
A:Introns: 19/2

C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match 4.1%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
|||||
Db 50 AAAAAAL 56

RESULT 19

US0705
antifreeze protein (clones 1A-1a and 3-3a) - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 24-Oct-2000
C:Accession: J50705
R: Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:9220995; PMID:155765
A:Accession: J50705
A:Molecule type: DNA
A:Residues: 1-82 <DNA>
A:Cross-references: GB:M62413; NID:g213586; PIDN:AAA49468.1; PID:g213587
C:Genetics:
A:Introns: 19/2
A:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match 4.1%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
|||||
Db 50 AAAAAAL 56

RESULT 20

151125
antifreeze protein - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Oct-2000
C:Accession: 151125
R: Pickett, M., Scott, G., Davies, P., Wang, N., Joshi, S., Hew, C.
Eur. J. Biochem. 143, 35-38, 1984
A:Title: Sequence of an antifreeze protein precursor.
A:Reference number: 151125; MUID:84285392; PMID:6547905
A:Accession: 151125
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-82 <PIC>
A:Cross-references: GB:M8337; NID:g213581; PIDN:AAA49466.1; PID:g213582
C:Genetics:
A:Gene: AFP
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match 4.1%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
|||||
Db 50 AAAAAAL 56

RESULT 21

A55846
OPPF homolog, kbmb 5'-region - Streptomyces tenebrarius (fragment)
C:Species: Streptomyces tenebrarius

C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 05-May-1995
C:Accession: A55846
R: Holmes, D.J.; Cundliffe, E.

Mol. Gen. Genet. 229, 229-237, 1991
A:Title: Analysis of a ribosomal RNA methylase gene from Streptomyces tenebrarius whi
A:Reference number: S17717; MUID:9201656; PMID:1921972
A:Accession: A55846
A:Molecule type: DNA
A:Residues: 1-87 <HOL>
A:Note: The authors neither translated nor discussed this partial open reading frame

Query Match 4.1%; Score 7; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 RRAVAVR 129
|||||
Db 36 RRAVAVR 42

RESULT 22

F86299
hypothetical protein F309.23 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: F86299
R: Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: F86299
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-150 <STO>
A:Cross-references: GB:AE005172; NID:g496364; PIDN:AAD34695.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 4.1%; Score 7; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
|||||
Db 17 AAAAAAL 23

RESULT 23

B70533
hypothetical protein Rv2719c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70533
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70533
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-165 <COL>
A:Cross-references: GB:Z96072; GB:AL123456; NID:g3261793; PIDN:CAB09460.1; PID:g21820
A:Experimental source: strain H37RV

C:Genetics:
A:Gene: RV2719c

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 165;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 VAPNAPT 39
DB 130 VAPNAPT 136

RESULT 24

S58217

outer membrane protein - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 26-Aug-1999

C:Accession: S58217

R:Lim, A.; de Vos, D.; Brauns, M.; Gaballa, A.; Hamers, R.; Cornelis, P.

Submitted to the EMBL Data Library, August 1995

A:Description: Molecular cloning, nucleotide sequence, and characterization of OprL, a

A:Reference number: S58216

A:Accession: S58217

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-167 <Lim>

A:Cross-references: EMBL:250191

C:Superfamily: outer membrane protein A

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 167;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 GHTDERG 112
DB 102 GHTDERG 108

RESULT 25

G83525

outer membrane protein OprL precursor PA0973 [imported] - Pseudomonas aeruginosa (strain

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G83525

R:Stover, C.K.; Pham, X.Q.; Ewitt, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; Lim,

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: G83525

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-168 <STO>

A:Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AA04362.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: oprL, PA0973

C:Superfamily: outer membrane protein A

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 168;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 GHTDERG 112
DB 102 GHTDERG 108

RESULT 26

I40346

ompL6 protein - Brucella abortus

C:Species: Brucella abortus

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999

C:Accession: I40346

R:Tibor, A.; Weynants, V.; Denoel, P.; Lichtfouse, B.; De Bolle, X.; Saman, E.; Limet

Infect. Immun. 62, 3633-3639, 1994

A:Title: Molecular cloning, nucleotide sequence, and occurrence of a 16.5-kilodalton

A:Reference number: I40346; MUID:94341863; PMID:8063379

A:Accession: I40346

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-168 <RES>

A:Cross-references: GB:L27996; NID:g619644; PIDN:AA59360.1; PID:g538293

C:Genetics:

A:Gene: pal

C:Superfamily: outer membrane protein A

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 168;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 WSONRRA 168
DB 153 WSONRRA 159

RESULT 27

AF3294

peptidoglycan-associated lipoprotein [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Mar-2002

C:Accession: AF3294

R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit

A:Reference number: AD3252; PMID:11756688

A:Accession: AF3294

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-168 <KUN>

A:Cross-references: GB:AE008917; PIDN:AA151521.1; PID:g17982237; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEI0340

A:Map position: I

C:Superfamily: outer membrane protein A

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 168;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 WSONRRA 168
DB 153 WSONRRA 159

RESULT 28

AE2405

hypothetical protein al14797 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AE2405

R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Irigu

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AE2405

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-170 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA076496.1; PID:g17133934; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Gene: all4797

Query Match 4.1%; Score 7; DB 2; Length 170;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 TGVAPLV 53
|||||
Db 58 TGVAPLV 64

RESULT 29

peptidoglycan-associated lipoprotein precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli
C:Date: 30-Sep-1988 #sequence-revision 30-Sep-1988 #text-change 01-Mar-2002
C:Accession: A27534; S20547; D64810

R:Chen, R.; Henning, U.
Eur. J. Biochem. 163, 73-77, 1987

A:Title: Nucleotide sequence of the gene for the peptidoglycan-associated lipoprotein of

A:Reference number: A27534; MUID:8713578; PMID:3545827

A:Accession: A27534

A:Molecule type: DNA

A:Cross-references: GB:X05123; NID:g42256; PIDN:CA428771.1; PID:g42257

R:Laazaroni, J.C.; Portaiter, R.
Mol. Microbiol. 6, 735-742, 1992

A:Title: The excC gene of Escherichia coli K-12 required for cell envelope integrity and

A:Reference number: S20546; MUID:92244043; PMID:1574003

A:Accession: S20547

A:Molecule type: DNA

A:Cross-references: EMBL:X65796; NID:g41358; PIDN:CA46673.1; PID:g41360

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.

.A.; Rose, D.J.; Mau, B.; Shaoh, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: D64810

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-173 <BLAT>

A:Cross-references: GB:A000177; GB:U00096; NID:g1786955; PIDN:AACT3835.1; PID:g1786962;

A:Experimental source: strain K-12, substrain MG1655

C:Comment: This lipoprotein of unknown function is very strongly associated with the per

C:Genetics:

A:Gene: pal; excC

A:Map position: 17 min

C:Superfamily: outer membrane protein A

C:Keywords: lipid binding; lipoprotein

F.1-21/Domains: signal sequence #status predicted <SIG>

F.22-173/Product: peptidoglycan-associated lipoprotein #status predicted <MAT>

Query Match 4.1%; Score 7; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 SLGERRA 125
|||||
Db 120 SLGERRA 126

RESULT 30

peptidoglycan-associated lipoprotein [imported] - Escherichia coli (strain O157:H7, subs

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 01-Mar-2002

C:Accession: A85577

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: A85577

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-173 <STO>

A:Cross-references: GB:A000174; NID:g12513674; PIDN:ANG5077.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: pal

C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 SLGERRA 125
|||||
Db 120 SLGERRA 126

RESULT 31

peptidoglycan-associated lipoprotein [imported] - Escherichia coli (strain O157:H7, s

C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence-revision 18-Jul-2001 #text-change 01-Mar-2002

C:Accession: H90725

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H90725

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-173 <MAT>

A:Cross-references: GB:BA000007; PIDN:BA34199.1; PID:g13360235; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: Ecs0776

C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 SLGERRA 125
|||||
Db 120 SLGERRA 126

RESULT 32

peptidoglycan-associated lipoprotein precursor [imported] - Salmonella enterica subsp

C:Species: Salmonella enterica subsp. enterica serovar Typh

A:Note: this species has also been called Salmonella typh

C:Date: 09-Nov-2001 #sequence-revision 09-Nov-2001 #text-change 01-Mar-2002

C:Accession: A10592

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, K.M.; Dowd, L.; White, N.; Farr

S.; Moutle, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; PMID:11677608

A:Accession: A10592

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-174 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05211.1; PID:g16501981; GSPDB:GN00176

C:Genetics:

A:Gene: STY0795

C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SLGERA 125
Db 121 SLGERA 127

RESULT 33

F72550

hypothetical protein APE1690 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: F72550

R:Kawabadyai, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: F72550

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-174 <KAM>

A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BA80691.1; PID:95105378

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1690

C:Superfamily: Aeropyrum pernix hypothetical protein APE1690

Query Match 4.1%; Score 7; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAAAAA 13
Db 89 TAAAAA 95

RESULT 34

T52661

cysteine synthase (EC 4.2.99.8), mitochondrial [imported] - Arabidopsis thaliana (fragma

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 03-Nov-2000

C:Accession: T52661

R:Hatzfeld, Y.; Maruyama, A.; Schmidt, A.; Noji, M.; Ishizawa, K.; Saito, K.

A>Title: Beta-cyanoalanine synthase is a mitochondrial cysteine synthase-like protein in

A:Reference number: 226167

A:Accession: T52661

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-176 <HAT>

A:Cross-references: EMBL:AF011044; PIDN:CA55622.1

A:Experimental source: cultivar Columbia

C:Genetics:

A:Gene: oas6

C:Superfamily: threonine dehydratase

C:Keywords: carbon-oxygen lyase; mitochondrion

Query Match 4.1%; Score 7; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 127 AAAAAA 133

RESULT 35

AE3013

omp16 protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Mar-2002

C:Accession: AE3013

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; MCCI

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AE3013

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-177 <KUR>

A:Cross-references: GB:AE00689; PIDN:AAU44523.1; PID:917742135; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: palA

A:Map position: linear chromosome

C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 WSONRRA 168
Db 162 WSONRRA 168

RESULT 36

B98271

omp16 protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)

C:Species: Agrobacterium tumefaciens

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 01-Mar-2002

C:Accession: B98271

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm

A.; Liu, F.; Mollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,

Science 294, 2323-2328, 2001

A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A:Reference number: A97359; PMID:11743194

A:Accession: B98271

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-177 <KUR>

A:Cross-references: GB:AE007870; PIDN:AKK89692.1; PID:915159600; GSPDB:GN00170

C:Genetics:

A:Gene: AGK_L_2246

A:Map position: linear chromosome

C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 WSONRRA 168
Db 162 WSONRRA 168

RESULT 37

D75351

Probable acetyltransferase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: D75351

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036966; PMID:10567266

A:Accession: D75351

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-179 <NH>
 A:Cross-references: GB:AE002021; GB:AE000513; NID:g6459573; PIDN:AAF11356.1; PID:g645957
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DRL800
 A:Map position: 1
 C:Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rimJ

Query Match 4.1%; Score 7; DB 2; Length 179;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AAAAAA 15
 |||||
 DB 108 AAAAAA 114

RESULT 38
 A82752
 hypothetical protein XF0883 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 16-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: A82752
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82752
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-185 <STM>
 A:Cross-references: GB:AE003927; GB:AE003849; NID:g9105783; PIDN:AAF83693.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kempfer, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF0883

Query Match 4.1%; Score 7; DB 2; Length 185;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
 |||||
 DB 93 AAAAAA 99

RESULT 39
 E71495
 probable peptidoglycan-associated lipoprotein - Chlamydia trachomatis (serotype D, strai
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 17-Mar-2000
 C:Accession: E71495
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
 A:Reference number: A71570; MUID:9900809; PMID:9784136
 A:Accession: E71495

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <ARN>
 A:Cross-references: GB:AE001330; GB:AE001273; NID:g3329034; PIDN:AAC68202.1; PID:g332
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: pal
 C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 188;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 GHTDERG 112
 |||||
 DB 119 GHTDERG 125

RESULT 40
 H72036
 peptidoglycan-associated lipoprotein CP1091 [imported] - Chlamydia pneumoniae (st
 C:Species: Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 19-May-2000
 C:Accession: H72036; C81504
 R:Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606; PMID:10192388
 A:Accession: H72036
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-192 <ARN>
 A:Cross-references: GB:AE001659; GB:AE001363; NID:g4377081; PIDN:AAD18919.1; PID:g437
 A:Experimental source: strain CWD029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 C.; Dodson, R.; Gilm, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150235; PMID:10664935
 A:Accession: C81504
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-192 <RNA>
 A:Cross-references: GB:AE002265; GB:AE002161; NID:g7189999; PIDN:AAF38862.1; PID:g719
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: pal; CP1091
 C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 192;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 GHTDERG 112
 |||||
 DB 123 GHTDERG 129

Search completed: July 6, 2003, 14:22:04
 Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 13:11:10 ; Search time 61 seconds
(without alignments)
375.723 Million cell updates/sec

Title: US-09-674-779b-2
Perfect score: 861
Sequence: 1 MMLHIQIAAAALSVLFM.....IARGTNEAMSNRAELSY 172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	861	100.0	172	21	AAV55089
2	855	99.3	172	21	AAV55090
3	850	98.7	172	21	AAV55091
4	844	98.0	172	21	AAV55092
5	266	30.9	153	11	AAV55092
6	264	30.7	153	9	AAV55093
7	264	30.7	153	9	AAV55094
8	264	30.7	153	11	AAV55095
9	264	30.7	153	15	AAV55161
10	259	30.1	153	9	AAV55161

11	215.5	25.0	179	18	AAW20394	H. pylori outer me
12	215.5	25.0	179	18	AAW20394	H. pylori outer me
13	215.5	25.0	179	18	AAW20394	H. pylori outer me
14	215.5	25.0	179	18	AAW20394	H. pylori outer me
15	215.5	25.0	179	18	AAW20394	H. pylori outer me
16	211.5	24.6	179	18	AAW20394	H. pylori outer me
17	208.5	24.2	179	18	AAW20394	H. pylori outer me
18	205.5	23.9	144	18	AAW20394	H. pylori outer me
19	204.5	23.8	192	22	AAW20394	H. pylori outer me
20	204.5	23.8	192	22	AAW20394	H. pylori outer me
21	204.5	23.8	192	22	AAW20394	H. pylori outer me
22	197.5	22.9	104	20	AAW20394	H. pylori outer me
23	196.5	22.8	214	20	AAW20394	H. pylori outer me
24	180.5	21.0	224	20	AAW20394	H. pylori outer me
25	176.5	20.5	353	22	AAW20394	H. pylori outer me
26	175.5	20.4	214	17	AAW20394	H. pylori outer me
27	174.5	20.3	338	16	AAW20394	H. pylori outer me
28	174	20.2	326	22	AAW20394	H. pylori outer me
29	171.5	19.9	161	17	AAW20394	H. pylori outer me
30	171.5	19.9	226	17	AAW20394	H. pylori outer me
31	171.5	19.9	350	22	AAW20394	H. pylori outer me
32	171.5	19.9	350	22	AAW20394	H. pylori outer me
33	170	19.7	369	21	AAW20394	H. pylori outer me
34	170	19.7	369	21	AAW20394	H. pylori outer me
35	169	19.6	359	21	AAW20394	H. pylori outer me
36	165.5	19.2	359	16	AAW20394	H. pylori outer me
37	165	19.2	228	23	AAW20394	H. pylori outer me
38	161.5	18.8	344	23	AAW20394	H. pylori outer me
39	161.5	18.8	344	23	AAW20394	H. pylori outer me
40	161.5	18.8	344	23	AAW20394	H. pylori outer me
41	161.5	18.8	344	23	AAW20394	H. pylori outer me
42	160	18.6	326	22	AAW20394	H. pylori outer me
43	158.5	18.4	344	23	AAW20394	H. pylori outer me
44	158.5	18.4	344	23	AAW20394	H. pylori outer me
45	158.5	18.4	344	23	AAW20394	H. pylori outer me
46	158.5	18.4	344	23	AAW20394	H. pylori outer me
47	155.5	18.1	341	23	AAW20394	H. pylori outer me
48	155.5	18.1	341	23	AAW20394	H. pylori outer me
49	155.5	18.1	341	23	AAW20394	H. pylori outer me
50	155.5	18.1	341	23	AAW20394	H. pylori outer me
51	150.5	17.5	344	23	AAW20394	H. pylori outer me
52	147	17.1	396	22	AAW20394	H. pylori outer me
53	145	16.8	287	22	AAW20394	H. pylori outer me
54	145	16.8	287	22	AAW20394	H. pylori outer me
55	144	16.7	230	20	AAW20394	H. pylori outer me
56	142.5	16.6	364	21	AAW20394	H. pylori outer me
57	142	16.5	672	20	AAW20394	H. pylori outer me
58	142	16.5	672	20	AAW20394	H. pylori outer me
59	138.5	16.1	679	20	AAW20394	H. pylori outer me
60	138.5	16.1	679	20	AAW20394	H. pylori outer me
61	136.5	15.9	137	23	AAW20394	H. pylori outer me
62	136.5	15.9	197	23	AAW20394	H. pylori outer me
63	136.5	15.9	335	16	AAW20394	H. pylori outer me
64	136.5	15.9	335	16	AAW20394	H. pylori outer me
65	136.5	15.9	335	16	AAW20394	H. pylori outer me
66	136.5	15.9	335	16	AAW20394	H. pylori outer me
67	136.5	15.9	344	18	AAW20394	H. pylori outer me
68	136.5	15.9	344	18	AAW20394	H. pylori outer me
69	136.5	15.9	344	18	AAW20394	H. pylori outer me
70	136.5	15.9	344	18	AAW20394	H. pylori outer me
71	136.5	15.9	344	18	AAW20394	H. pylori outer me
72	136.5	15.9	344	18	AAW20394	H. pylori outer me
73	136.5	15.9	344	18	AAW20394	H. pylori outer me
74	136.5	15.9	344	18	AAW20394	H. pylori outer me
75	136.5	15.9	344	18	AAW20394	H. pylori outer me
76	136.5	15.9	344	18	AAW20394	H. pylori outer me
77	136.5	15.9	344	18	AAW20394	H. pylori outer me
78	136.5	15.9	344	18	AAW20394	H. pylori outer me
79	136.5	15.9	344	18	AAW20394	H. pylori outer me
80	136.5	15.9	344	18	AAW20394	H. pylori outer me
81	136	15.8	452	22	AAW20394	H. pylori outer me
82	130.5	15.2	387	22	AAW20394	H. pylori outer me
83	122.5	14.2	550	22	AAW20394	H. pylori outer me

84	122.5	14.2	550	22	ABB52693	Escherichia coli p
85	107.5	12.5	390	23	AAU09397	Soluble Porphyrom
86	107.5	12.5	391	20	AAV34490	Porphyromonas ging
87	107.5	12.5	395	20	AAV34365	Porphyromonas ging
88	107	12.4	380	23	AAU09398	Soluble Porphyrom
89	107	12.4	385	20	AAV34491	Porphyromonas ging
90	107	12.4	387	20	AAV34366	Porphyromonas ging
91	99.5	11.6	196	22	AAU03573	P. gingivalis chim
92	99.5	11.6	315	22	ABG17777	Novel human diagno
93	96	11.1	368	20	AAV19945	B. burgdorferi ant
94	96	11.1	389	20	AAV19944	Hepatitis GB virus
95	87.5	10.2	507	16	AAH82451	Hepatitis GB virus
96	87.5	10.2	507	21	AAH809475	Hepatitis GB virus
97	87.5	10.2	1422	16	AAH82068	Hepatitis GB virus
98	87.5	10.2	1422	21	AAH809038	Hepatitis GB virus
99	87.5	10.2	2864	16	AAH82072	Hepatitis GB virus
100	87.5	10.2	2864	21	AAH809268	Hepatitis GB virus

ALIGNMENTS

RESULT 1
ID AAV55089 standard; Protein: 172 AA.

AAV55089;

01-MAR-2000 (first entry)

M. catarrhalis BASB019 protein sequence #1.

BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
genetic mutation screening; antibody production; vaccine; otitis media;
bacterial infection; pneumonia; sinusitis; nosocomial infection;
invasive disease; delayed speech learning; bacteria adhesion prevention;
upper respiratory tract infection; middle ear infection; therapy.

Moraxella catarrhalis.

WO9957277-A2.

11-NOV-1999.

03-MAY-1999; 99WO-EP03038.

06-MAY-1998; 98GB-0009683.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-062148/05.

N-PSDB: AA240351.

Novel BASB019 polynucleotides and polypeptides from Moraxella

catarrhalis used to prepare vaccines against bacterial infections

Claim 3; Fig 3; 101pp: English.

This sequence is a Moraxella catarrhalis BASB019 protein of the
invention. The sequences can be used for diagnosis of disease, staging of
disease, or determining response of an infectious organism to drugs. The
polynucleotides may be used as a source for hybridisation probes, and for
screening of genetic mutations, serotype, organism or strain
identification. Identification of mutation in BASB013 sequences, and as
components of arrays which are useful for diagnostic and prognostic
purposes. The polypeptides can be used to produce antibodies, and as a
target for the screening of antimicrobial drugs. The polypeptides can
also be used in vaccine formulations, and to identify agonists and
antagonists. The polypeptides, antibodies, agonists and antagonists
(which are bacteriostatic) are used for the treatment and prevention of
diseases including bacterial infection, otitis media in infants and

CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of in-dwelling devices or by other surgical techniques.
CC The frequency of Moraxella catarrhalis infections has risen dramatically,
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.

Sequence 172 AA;
Query Match 100.0%; Score 861; DB 21; Length 172;
Best Local Similarity 100.0%; Pred. No. 8; e-86;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMHIOIAAAALSVTFMTGCANKSTGVVAVAPPTGYTGVYTGAVLVNDENVK 60
DB 1 MMHIOIAAAALSVTFMTGCANKSTGVVAVAPPTGYTGVYTGAVLVNDENVK 60
QY 61 ALASKLPSTLYVDFEDSDDEIKPQAAALILDEQAFLETTOTARYLVAGHTDERGSEYKNSL 120
DB 61 ALASKLPSTLYVDFEDSDDEIKPQAAALILDEQAFLETTOTARYLVAGHTDERGSEYKNSL 120
QY 121 GERRAVAVRNYLLGKGINQASVELISFGEERPIAGTNEANSQRRELISY 172
DB 121 GERRAVAVRNYLLGKGINQASVELISFGEERPIAGTNEANSQRRELISY 172

RESULT 2

AAV55090 standard; Protein: 172 AA.

AAV55090;

01-MAR-2000 (first entry)

M. catarrhalis BASB019 protein sequence #2.

BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
genetic mutation screening; antibody production; vaccine; otitis media;
bacterial infection; pneumonia; sinusitis; nosocomial infection;
invasive disease; delayed speech learning; bacteria adhesion prevention;
upper respiratory tract infection; middle ear infection; therapy.

Moraxella catarrhalis.

WO9957277-A2.

11-NOV-1999.

03-MAY-1999; 99WO-EP03038.

06-MAY-1998; 98GB-0009683.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-062148/05.

N-PSDB: AA240352.

Novel BASB019 polynucleotides and polypeptides from Moraxella

catarrhalis used to prepare vaccines against bacterial infections

Claim 3; Fig 3; 101pp: English.

This sequence is a Moraxella catarrhalis BASB019 protein of the

invention. The sequences can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridisation probes, and for screening of genetic mutations, serotype, organism or strain identification, identification of mutation in BASB019 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies, and as a target for the screening of antimicrobial drugs. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases including bacterial infection, otitis media in infants and children, pneumonia in the elderly, sinusitis, nosocomial infections and invasive diseases, chronic otitis media with hearing loss, fluid accumulation in the middle ear, auditive nerve damage, delayed speech learning, infection of the upper respiratory tract and middle ear infection. They are also used in the prevention of adhesion of bacteria to eukaryotic matrix proteins on in-dwelling devices, or to extracellular proteins on wounds, and to thus prevent tissue damage and/or block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques, and it is no longer common to isolate *M. catarrhalis* strains that are resistant to standard antibiotics. The BASB019 products of the invention can be used screen for new antibacterial compounds that may target these resistant bacteria.

Sequence 172 AA;

Query Match 99.3%; Score 85; DB 21; Length 172;
Best Local Similarity 99.4%; Pred. No. 4e-85;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MMLHIQIAAAALSVLFTMTGCAKSTSQVAVAPNPTGYTYGVAPLVNDDETVK 60
Db 1 MMLHIQIAAAALSVLFTMTGCAKSTSQVAVAPNPTGYTYGVAPLVNDDETVK 60
QY 61 ALASKPLSLVYFDPDSDEIKPOAAALIDEOAQFLTTNQTARVLYAGHTDGRSGREYNMSL 120
Db 61 ALASTPLSLVYFDPDSDEIKPOAAALIDEOAQFLTTNQTARVLYAGHTDGRSGREYNMSL 120
QY 121 GERAAVAVRNLYLLGKINGQASVEIISFGERPFIAGTNEAMSONRRALSY 172
Db 121 GERAAVAVRNLYLLGKINGQASVEIISFGERPFIAGTNEAMSONRRALSY 172

RESULT 3

AAVS5091
ID AAVS5091 standard; Protein; 172 AA.

AC AAVS5091;
DT 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein sequence #3.

KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
KW upper respiratory tract infection; middle ear infection; therapy.

XX Moraxella catarrhalis.

XX WO957277-A2.

XX 11-NOV-1999.

XX 03-MAY-1999; 99WO-EP03038.

XX 06-MAY-1998; 98GB-0009683.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J;
XX WPI: 2000-062148/05.
XX N-PSDB: AA240353.

PT Novel BASB019 polynucleotides and polypeptides from Moraxella
catarrhalis used to prepare vaccines against bacterial infections
PS Claim 3; Fig 3; 101pp; English.

This sequence is a Moraxella catarrhalis BASB019 protein of the invention. The sequences can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridisation probes, and for screening of genetic mutations, serotype, organism or strain identification, identification of mutation in BASB019 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies, and as a target for the screening of antimicrobial drugs. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases including bacterial infection, otitis media in infants and children, pneumonia in the elderly, sinusitis, nosocomial infections and invasive diseases, chronic otitis media with hearing loss, fluid accumulation in the middle ear, auditive nerve damage, delayed speech learning, infection of the upper respiratory tract and middle ear infection. They are also used in the prevention of adhesion of bacteria to eukaryotic matrix proteins on in-dwelling devices, or to extracellular proteins on wounds, and to thus prevent tissue damage and/or block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques, and it is no longer common to isolate *M. catarrhalis* strains that are resistant to standard antibiotics. The BASB019 products of the invention can be used screen for new antibacterial compounds that may target these resistant bacteria.

Sequence 172 AA;

Query Match 98.7%; Score 85; DB 21; Length 172;
Best Local Similarity 98.3%; Pred. No. 1.4e-84;
Matches 169; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMLHIQIAAAALSVLFTMTGCAKSTSQVAVAPNPTGYTYGVAPLVNDDETVK 60
Db 1 MMLHIQIAAAALSVLFTMTGCAKSTSQVAVAPNPTGYTYGVAPLVNDDETVK 60

QY 61 ALASKPLSLVYFDPDSDEIKPOAAALIDEOAQFLTTNQTARVLYAGHTDGRSGREYNMSL 120
Db 61 TLASTPLSLVYFDPDSDEIKPOAAALIDEOAQFLTTNQTARVLYAGHTDGRSGREYNMSL 120

QY 121 GERAAVAVRNLYLLGKINGQASVEIISFGERPFIAGTNEAMSONRRALSY 172
Db 121 GERAAVAVRNLYLLGKINGQASVEIISFGERPFIAGTNEAMSONRRALSY 172

RESULT 4

AAVS5092
ID AAVS5092 standard; Protein; 172 AA.

XX AAVS5092;

XX 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein sequence #4.

KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
KW upper respiratory tract infection; middle ear infection; therapy.

XX Moraxella catarrhalis.
 OS W09957277-A2.
 PN 11-NOV-1999.
 PD 03-MAY-1999; 99WO-EP03038.
 PE 06-MAY-1998; 98GB-0009683.
 PR (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Ruelle J;
 XX WPI: 2000-062148/05.
 DR N-PSDB; AA240354.
 XX
 PT Novel BASB019 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections
 PS
 Claim 3; Fig 3; 101pp; English.

CC This sequence is a Moraxella catarrhalis BASB019 protein of the
 CC invention. The sequences can be used for diagnosis of disease, staging of
 CC disease, or determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and for
 CC screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies, and as a
 CC target for the screening of antimicrobial drugs. The polypeptides can
 CC also be used in vaccine formulations, and to identify agonists and
 CC antagonists. The polypeptides, antibodies, agonists and antagonists
 CC (which are bacteriostatic) are used for the treatment and prevention of
 CC diseases including bacterial infection, otitis media in infants and
 CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed speech
 CC learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarrhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarrhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used to screen for new antibacterial compounds that may target these
 CC resistant bacteria.

SO Sequence 172 AA;

Query Match 98.0%; Score 844; DB 21; Length 172;

Best Local Similarity 98.3%; Pred. No. 6, 3e-84;
 Matches 169; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MMLHIQIAAAAAALSVLFMTGCAKSTSQVAVAPNAPFGYTGVIYTVGAPLVNDDETVK 60
 DB 1 MMLHIQIAAAAAALSVLFMTGCAKSTSQVAVAPNAPFGYTGVIYTVGAPLVNDDETVK 60
 QY 61 ALASKLPVLVYFDFSDDEIKPOAAALIDEOAQLTTNQTARVLVAGHDERGSEYNNSL 120
 DB 61 ALASTLPVLVYFDFSDDEIKPOAAALIDEOAQLTTNQTARVLVAGHDERGSEYNNSL 120
 QY 121 GERRAVAVRNYLLGKGINQASVELISFGEERPIAGTNEAASQRRRELST 172
 DB 121 GERRAVAVRNYLLGKGINQASVELISFGEERPIAGTNEAASQRRRELST 172

RESULT 5
 AAR07145
 ID AAR07145 standard; protein: 153 AA.

XX AAR07145;

AC 24-JAN-1991 (first entry)

DE 16.6KD outer membrane protein (OMP) of H. influenzae.

KW HI vaccine; OMP.

OS Haemophilus influenzae.

PN EP389925-A.

PD 03-OCT-1990.

PE 20-MAR-1990; 90EP-0105205.

PR 29-MAR-1989; 89US-0330229.

PA (UYNE-) STATE UNIV NEW YORK.

PI Murphy TF, Apicella MA;

DR WPI: 1990-298924/40.

DR N-PSDB; AAR07145.

PT Purification of outer membrane protein of haemophilus influenzae
 PT - by separation from an insoluble fraction using
 PT detergent-contg., then detergent-free buffers.

PS Disclosure; Page 8; 22pp; English.

CC Method claimed produces large quantities of the purified OMP, useful
 CC in raising antibodies for detection, and as a vaccine against
 CC H. influenzae.

SO Sequence 153 AA;

Query Match 30.9%; Score 266; DB 11; Length 153;

Best Local Similarity 36.7%; Pred. NO. 6, 5e-21;
 Matches 62; Conservative 24; Mismatches 45; Indels 38; Gaps 3;

QY 25 NKSTSQVAVAPNAPF-----GYTGVIYTVGAPLVNDDETVKALA 63
 DB 2 NKFKSLVAVASVAPALAAACSSNNDAAGNAGAAOFGGYS-----VADLQORNT----- 50
 QY 64 SKLPVLVYFDFSDDEIKPOAAALIDEOAQLTTNQTARVLVAGHDERGSEYNNSLGGR 123
 DB 51 -----VYFGPDKDYDITGEYVQILDAAHAAYINATPAKVLVSGTDERGTPPEYNIALGGR 104
 QY 124 RAVAVRNYLLGKGINQASVELISFGEERPIAGTNEAASQRRRELST 172
 DB 105 RADAVKGYLAGKGVAGKLGTVSYGGEERPAVLGHDEAAYSQRRRAVLAY 153

RESULT 6
 AAR82947
 ID AAR82947 standard; protein: 153 AA.

AC AAR82947;

DE 10-OCT-1990 (first entry)

DE 16600 dalton outer membrane protein of non-typable H influenzae.

KW Haemophilus influenzae; 16600 dalton outer membrane protein (OMP);
 KW pneumonia; meningitis; ss.

OS Haemophilus influenzae.

PN EP281673-A.

PD 14-SEP-1988.

PT Plasmids, E.coli, hybridomas and antibodies from H.influenzae protein
PT for use as vaccine and detection of Haemophilus influenzae

The PBOM proteins were isolated from a PBOM-enriched insoluble cell wall fraction from physically disrupted cells of *H. influenzae* and then sublocalising the PBOM from the cell wall fraction by heating in the presence of a detergent or digesting the cell wall fraction with lysozyme, opt. in the presence of a detergent. The genes encoding the PBOM proteins were isolated by screening a DNA library with an oligonucleotide probe based on the amino acid sequence of the PBOM protein, or using antibodies to PBOM.

Query Match	30.7%;	Score 264;	DB 11;	Length 153;
Best Local Similarity	49.5%;	Pred. No. 1.1e-20;		

DT 29-JUL-1997 (first entry)
 XX
 DE H. pylori outer membrane protein 31262.aa.
 XX
 KW Outer membrane; vaccine; prevention; treatment; infection; envelope;
 XX identification; binding compound; bacterium; life cycle; activator;
 KW bacterium; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PN MO9640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaard BL;
 XX
 DR WPI: 1997-052306/05.
 DR N-PSDB; AAT67789.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 56; Page 577; 1481pp: English.
 XX
 CC This sequence is a H. pylori outer membrane protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC Useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 CC
 SO Sequence 179 AA;
 XX
 Query Match 25.0%; Score 215.5; DB 18; Length 179;
 Best Local Similarity 32.8%; Pred. No. 2.6e-15;
 Matches 57; Conservative 30; Mismatches 66; Indels 21; Gaps 5;
 XX
 OY 15 SVLTFFM-----TGCANR-----STSOVMVAP--NAPGTGYTYGVAPLVND 56
 XX
 DB 5 SVSFSLVAFLLVVGCSHKMDNKTVAAGVSTKAVGTAPVTEPAPEKPEKQEPAPVVEEK 64
 XX
 OY 57 EYKALASKLPISLVYFDEPDEIKPQAAAILDEQAQFLTTQTAQVAVAGHTDEGRSREY 116
 XX
 DB 65 PAIE--SGTIIASITFDPKYEIKSDDETLDEIVQAKENH-MQVLLGNTDEFGSSEY 121
 XX
 OY 117 NMSLGERRAVAVRNVLGKGINQASVELISFGEERPIAFGTNEBAMQONRAEL 170
 XX
 DB 122 NOALGVKRTLSVKNALVTKGVKDKIKTISFGESKPCYQKTRCTRYENRRADV 175
 XX
 RESULT 12
 AAM24651
 ID AAM24651 standard; Protein; 179 AA.
 XX
 AC AAM24651;
 XX

DT 11-AUG-1997 (first entry)
 XX
 DE H. pylori outer membrane protein 31262.aa.
 XX
 KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
 KW activator; inhibitor; bacterial life cycle; vaccine; immune;
 KW detection; antisense; inhibition.
 XX
 OS Helicobacter pylori.
 XX
 PN MO9719098-A1.
 XX
 PD 29-MAY-1997.
 XX
 PF 15-NOV-1996; 96WO-US18542.
 XX
 PR 17-NOV-1995; 95US-0561469.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Smith DH;
 XX
 DR WPI: 1997-298052/27.
 DR N-PSDB; AAT77469.
 XX
 PT Helicobacter pylori nucleic acid sequences and related proteins -
 PT used for diagnostics and therapeutics
 XX
 PS Claim 18; Page 171; 235pp: English.
 XX
 CC This sequence represents an H. pylori outer membrane protein.
 CC Helicobacter pylori has been strongly linked to chronic gastritis and
 CC duodenal ulcer disease. The nucleic acid sequences of the invention
 CC are used to evaluate compounds, especially activators or inhibitors of
 CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
 CC sequence. The nucleic acid sequences, and corresponding proteins, are
 CC also useful for generating vaccines for immunising subjects against H.
 CC pylori or for use in detecting the presence of Helicobacter species in
 CC a sample. Antisense nucleic acid sequences of these sequences are
 CC used to inhibit expression of a gene from Helicobacter species. H.
 CC pylori whole genomic DNA was isolated and nebulised to a median size of
 CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
 CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
 CC complementary to the BstXI-cut pmx vectors, while the overhang is not
 CC self-complementary. Therefore the linkers will not concatamerise nor
 CC will the cut vector re-ligate itself easily. The linker-adaptor inserts
 CC were ligated to each of the 20 pmx vectors to construct a series of
 CC shotgun subclone libraries. The purified DNA samples were then
 CC sequenced.
 CC Note: The ORF/protein reference number for this sequence was obtained
 CC from the related specification, WO9640893.
 CC
 SO Sequence 179 AA;
 XX
 Query Match 25.0%; Score 215.5; DB 18; Length 179;
 Best Local Similarity 32.8%; Pred. No. 2.6e-15;
 Matches 57; Conservative 30; Mismatches 66; Indels 21; Gaps 5;
 XX
 OY 15 SVLTFFM-----TGCANR-----STSOVMVAP--NAPGTGYTYGVAPLVND 56
 XX
 DB 5 SVSFSLVAFLLVVGCSHKMDNKTVAAGVSTKAVGTAPVTEPAPEKPEKQEPAPVVEEK 64
 XX
 OY 57 EYKALASKLPISLVYFDEPDEIKPQAAAILDEQAQFLTTQTAQVAVAGHTDEGRSREY 116
 XX
 DB 65 PAIE--SGTIIASITFDPKYEIKSDDETLDEIVQAKENH-MQVLLGNTDEFGSSEY 121
 XX
 OY 117 NMSLGERRAVAVRNVLGKGINQASVELISFGEERPIAFGTNEBAMQONRAEL 170
 XX
 DB 122 NOALGVKRTLSVKNALVTKGVKDKIKTISFGESKPCYQKTRCTRYENRRADV 175
 XX
 RESULT 13

ID AAM20795 standard; Protein; 187 AA.
 AC AAM20795;
 DT 21-JUL-1997 (first entry)
 DE H. pylori outer membrane protein 079p31516orf4.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN MO9640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; 96MO-US09122.
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 1997-052306/05.
 DR N-PSDB; AAT68048.
 XX
 XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 infection, to detect Helicobacter
 PS Claim 56; Page 1203; 1481pp; English.
 XX
 CC This sequence is a H. pylori outer membrane protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. No
 CC identity likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 187 AA;
 Query Match 25.0%; Score 215.5; DB 18; Length 187;
 Best Local Similarity 32.8%; Pred. No. 2,8e-15;
 Matches 57; Conservative 30; Mismatches 66; Indels 21; Gaps 5;
 QY 15 SVUTM-----TGCAK-----STQVAVP--NAPGTGYITGVAPLVNDND 56
 DB 13 SVESFLVAFLLVGGSHKMDKNTVAGDVSTKAVGTAPVTEPAPEKEEPKQEPAPVVEE 72
 QY 57 ETVKALASKLPISLVYFDDSDDEIKPQAAAILDEQAQFLITQTARVLYAGHTDERSGREY 116
 DB 73 PAIE--SGTITASTFEDDKYEIKESDQETLDELIVQAKENH-MQVLLLEANTDEFGSSSEY 129
 QY 117 NMSIGERRAVAVRNVLKGGINQASVETISFGEEPIAFGTNEBAMSONRAEL 170
 DB 130 NQALGVKRTLSVKNALVTKGVEKDMITISFGESKPCVCKTRECYNENRVDV 183

XX
 AC AAM89813;
 DT 18-FEB-1999 (first entry)
 DE Protein encoded by clone G1a ORF2.
 KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
 OS Helicobacter pylori.
 PN WO9849314-A2.
 PD 05-NOV-1998.
 PF 27-APR-1998; 98MO-US08487.
 PR 14-OCT-1997; 97US-0061958.
 PR 25-APR-1997; 97US-0045107.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Chow TP, Fry KE, Lim MY, McAttee CP;
 DR WPI: 1999-009433/01.
 DR N-PSDB; AAV90545.
 XX
 XX New Helicobacter pylori antigens and related nucleic acid sequences
 PT - useful in serological diagnosis and protective vaccines, providing
 PS long-lasting immune response
 PS Claim 10; Page 96-97; 402pp; English.
 XX
 CC The present sequence represents a Helicobacter pylori antigenic protein
 CC that is characterised by immunoreactivity with H. pylori positive
 CC antisera. The proteins are highly immunogenic and induce a long-lasting
 CC immune response that persists even after antimicrobial treatment. In
 CC antibody detection assays, on sera, plasma, urine, saliva etc., they are
 CC highly sensitive and specific. The specification also describes 69
 CC previously unrecognised immunogenic cluster families. H. pylori antigens
 CC are used to detect H. pylori-specific antibodies, for diagnosing
 CC infection or to confirm eradication of infection, and in vaccines to
 CC protect against H. pylori infection and related diseases (gastritis,
 CC peptic ulcer, gastric adenocarcinoma/lymphoma).
 XX
 SQ Sequence 179 AA;
 Query Match 24.6%; Score 211.5; DB 20; Length 179;
 Best Local Similarity 31.2%; Pred. No. 7,2e-15;
 Matches 54; Conservative 36; Mismatches 72; Indels 11; Gaps 4;
 QY 8 AAAAAALSVLTMTTCGCAKSTSQVAVP-NAPGTGYITGVAPLVNDDEPKA-LASK 65
 DB 4 SSAFEFLVAFLLVACGSHKMDKNTVAGDVSAKTQGTAPVTEPAPEKEEPKQEPAPVVEE 63
 QY 66 LPST-----VYDFPDSDDEIKPQAAAILDEQAQFLITQTARVLYAGHTDERSGREY 117
 DB 64 KRAISGTTIASIYDFPKYEIKESDQETLDELIVQAKENH-MQVLLLEANTDEFGSSSEY 122
 QY 118 MSLGERRAVAVRNVLKGGINQASVETISFGEEPIAFGTNEBAMSONRAEL 170
 DB 123 QALGVKRTLSVKNALVTKGVEKDMITISFGESKPCVCKTRECYNENRVDV 175

RESULT 16
 AAM89813
 ID AAM89813 standard; Protein; 179 AA.

RESULT 17
 AAM23592
 ID AAM23592 standard; Protein; 179 AA.

AC AAM23592;
 DT 23-JAN-1998 (first entry)

DE H. pylori Omp22 outer membrane protein.
 XX Omp22 gene; outer membrane protein; antigen; immunogen; stomach;
 KW gastritis; peptic ulcers; gastric cancer; diagnostic; vaccine.
 XX Helicobacter pylori.
 OS
 FH Key Location/Qualifiers
 FT Peptide 1..35
 FT /label= leader_sequence
 FT 36..179
 FT Protein /label= mature_peptide
 FT /note= "Omp22"
 XX
 XX WO9728264-A1.
 XX
 XX 07-AUG-1997.
 XX
 XX 06-SEP-1996; 96WO-KR00154.
 XX
 XX 30-JAN-1996; 96KR-0002105.
 XX
 XX (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX
 XX Kim J, Moon H, Park Y, Seo W, Yu G, Yum J;
 XX WPI: 1997-402617/37.
 XX N-PSDB: AAT74195.
 XX
 XX Helicobacter pylori outer membrane protein, Omp22 - used in a
 PT vaccine for treatment or prevention of H. pylori infection
 PS
 XX Example 3: Fig 5; 38pp; English.
 CC This sequence represents Omp22, an outer membrane protein which
 CC exhibits antigenicity and immunogenicity against Helicobacter pylori.
 CC This bacterium is associated with inflammation of the stomach and
 CC gastritis related diseases e.g. peptic ulcers and gastric cancer. Omp22
 CC could be used in immunological therapy as a H. pylori-specific antigen
 CC for the treatment and prevention of diseases associated with this
 CC microorganism e.g. as the active ingredient in a diagnostic kit or a
 CC vaccine.
 CC
 XX
 XX Sequence 179 AA:
 SQ
 Query Match 24.2%; Score 208.5; DB 18; Length 179;
 Best Local Similarity 30.8%; Pred. No. 1.5e-14;
 Matches 52; Conservative 33; Mismatches 69; Indels 15; Gaps 4;
 QY 14 LSVLFMTGCMKSTQVMVAF-NAPGTGVIYTVG-----APLVNDDETVKA 61
 DB 10 LVAFLFIAGCKHNDKREIVAGDVSAKAVQSPVSTELAQKQEPKQEPAPVVEKPAVE- 68
 QY 62 LASKLPISLVYDFDSDDEIKPQAAAIDDEQAQFLTTNOTARYLVAGHTDEGSEYKMSIG 121
 DB 69 -SGTIIASIVYDFDDEIKESDQETLDEIVQKAKENH-MQVLLGNTDEGSSSEYQALG 126
 QY 122 ERRNAVAVRNYLLGKINQASVEIISFGEPRIAGTNEEAMSONRAEL 170
 DB 127 VKRTLIVKNAIVIKGVERKDMTKTISFGETKPCQKTRCKEKERRVDV 175

DE gastritis; peptic ulcers; gastric cancer; diagnostic; vaccine;
 KW recombinant protein.
 XX Helicobacter pylori.
 OS
 FH Key Location/Qualifiers
 FT Peptide 1..144
 FT /label= Omp22
 FT Protein /note= "Recombinant outer membrane protein"
 XX
 XX WO9728264-A1.
 XX
 XX 07-AUG-1997.
 XX
 XX 06-SEP-1996; 96WO-KR00154.
 XX
 XX 30-JAN-1996; 96KR-0002105.
 XX
 XX (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX
 XX Kim J, Moon H, Park Y, Seo W, Yu G, Yum J;
 XX WPI: 1997-402617/37.
 XX N-PSDB: AAT74194.
 XX
 XX Helicobacter pylori outer membrane protein, Omp22 - used in a
 PT vaccine for treatment or prevention of H. pylori infection
 PS
 XX Claim 2: Pages 25; 38pp; English.
 CC This amino acid sequence of Omp22, is a recombinant outer membrane
 CC protein which is identical to the mature protein of its native form
 CC (see AAW23592). It retains antigenicity and immunogenicity against
 CC Helicobacter pylori. This recombinant Omp22 protein can be mass produced
 CC in via expression in E. coli DH5-alpha cells. H. pylori is associated
 CC with inflammation of the stomach and gastritis related diseases e.g.
 CC peptic ulcers and gastric cancer. The native and the recombinant form of
 CC Omp22 could be used in immunological therapy as a H. pylori-specific
 CC antigen for the treatment and prevention of diseases associated with
 CC this microorganism e.g. as the active ingredient in a diagnostic kit or
 CC a prophylactic/therapeutic vaccine
 CC
 XX
 XX Sequence 144 AA:
 SQ
 Query Match 23.9%; Score 205.5; DB 18; Length 144;
 Best Local Similarity 38.0%; Pred. No. 2.4e-14;
 Matches 46; Conservative 25; Mismatches 47; Indels 3; Gaps 2;
 QY 50 APLVNDDETVKALASKLPISLVYDFDSDDEIKPQAAAIDDEQAQFLTTNOTARYLVAGHTD 109
 DB 23 APVVEKPAVE--SGTIIASIVYDFDDEIKESDQETLDEIVQKAKENH-MQVLLGNTD 79
 QY 110 ERGSEYKMSIGERRAVAVRNYLLGKINQASVEIISFGEPRIAGTNEEAMSONRAEL 169
 DB 80 EFGSEYQALGVRRLISVKNALIVIKGVERKDMTKTISFGETKPCQKTRCKEKERRVDV 139
 QY 170 L 170
 DB 140 V 140

RESULT 18
 AAW23591
 ID AAW23591 standard; Protein; 144 AA.
 XX
 XX AAW23591;
 AC
 XX
 XX 23-JAN-1998 (first entry)
 DT
 XX H. pylori recombinant Omp22 recombinant outer membrane protein.
 DE
 XX Omp22 gene; outer membrane protein; antigen; immunogen; stomach;
 KW

RESULT 19
 AAY35412
 ID AAY35412 standard; Protein; 192 AA.
 XX
 XX AAY35412;
 AC
 XX
 XX 13-SEP-1999 (first entry)
 DT
 XX Chlamydia pneumoniae transmembrane protein sequence.
 DE
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW

20-DEC-2000: 2000W0-C201E34

X
D 10-JAN-2002.

X
D 10-JAN-2002.

XX Griffiths R;
XX WPI: 1999-371125/31.
XX
XX Genome sequence of Chlamydia trachomatis
XX
XX Disclosure; Page 1248-1249; 1755pp; English.
XX
XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
XX of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
XX vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
XX can also be used to control growth of the microorganism. Chlamydia
XX trachomatis is responsible for a large number of diseases, e.g. eye
XX diseases such as conventional trachoma, nonendemic trachoma,
XX paratrachoma, and inclusion conjunctivitis; genital diseases such as
XX nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
XX perlepetitis, bartonellosis, pneumopathy in breast feeding infants;
XX and venereal lymphogranulomatosis. The polypeptides of the invention
XX may be of use in treating these diseases.

SO Sequence 214 AA;

Query Match 22.8%; Score 196.5; DB 20; Length 214;
Best Local Similarity 35.3%; Pred. No. 4e-13;
Matches 49; Conservative 19; Mismatches 54; Indels 17; Gaps 2;

OY 48 GVAFLVNDDETVKALASKLPSTL-----VYEDPSDEIKRQ-AAATLDEQ 90
DB 70 GFVPEYSDDEIQAFVEDFSKEEQLYKTSAGSTSPFNITFADTSIKGEDNLTTLASL 129
OY 91 AOFLLTNOTAVLVAAGHTDERGSRFYMMSLGERRAVAARNYLLGKGINQASVEIISFCEE 150
DB 130 VRHHSKPKATLYIEGHTDERGAAYNALGARRAVAARNYLLGKGINQASVEIISFCEE 189
OY 151 RPIAFGTNEAASQNRRAE 169
DB 190 HPVHPGHNELAMOOONRRTE 208

RESULT 24

AAB20105
ID AAB20105 standard; Protein; 224 AA.

XX AAB20105;

XX 23-APR-2001 (first entry)

XX Moraxella catarrhalis BASB113 protein.

XX BASB113; infection; otitis media; pneumonia; therapy; diagnosis;

XX antibacterial; antimicrobial.

XX Moraxella catarrhalis.

XX WO200100836-A1.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-EP05851.

XX 25-JUN-1999; 99GB-0015044.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonnard J;

XX WPI: 2001-112458/12.

XX N-PSDB: AAF30043.

XX New BASB113 polypeptide isolated from Moraxella catarrhalis bacterium,
XX useful for diagnosing and producing vaccines against bacterial
XX infections such as otitis media and pneumonia

XX Claim 1; Page 67; 86pp; English.
PS
XX

The present sequence is that of BASB113 protein from Moraxella
catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
media in children and pneumonia in adults. The invention provides
BASB113 polypeptides, and polynucleotides encoding them, as well as
expression vectors, host cells and methods for producing BASB113
polypeptides using recombinant methods. Also claimed is a vaccine
composition comprising a BASB113 polypeptide, an immunogenic
fragment of a BASB113 polypeptide, or a polypeptide having at least
85% amino acid sequence identity to BASB113, or comprising a
polynucleotide encoding such a polypeptide. A claimed method of
diagnosing a Moraxella infection involves identifying a BASB113
polypeptide or antibody. A claimed therapeutic composition useful
in treating humans with M. catarrhalis infection comprises at least
1 antibody directed against a BASB113 polypeptide. BASB113
polypeptides also have utility in raising specific antibodies,
and in screening for antibacterial drugs.

SO Sequence 224 AA;

Query Match 21.0%; Score 180.5; DB 22; Length 224;
Best Local Similarity 27.0%; Pred. No. 2.4e-11;
Matches 48; Conservative 25; Mismatches 88; Indels 17; Gaps 3;

OY 7 IAAAAALSVLTFTMGANKSTSOVMYAPNAPGYGVY-----TGVAPL 52
DB 42 IGAVAGALGGTALSKANGKEKTRDAIL-GAAYGAAGAAMERQAKQIDQMOGGYGVY-- 98
OY 53 VDNDEYKALASKIPSLVYEDPSDEIKPQAAIIDEQAOFLTNOTAVLVAAGHTDERG 112
DB 99 VTHDTGNLNTLMPGNTFPAHDDTLNSAFLGRNLQANTMNOGHETTIVVGHITDSTG 158
OY 113 SREYNMSLGERRAVAARNYLLGKGINQASVEIISFCEEPIAFGTNEAASQNRRAEL 170
DB 159 QAAYNOELSRRAVRYVYLINQGVDPYRIQVYGVGMKPIASNATEGRQONRRVEL 216

RESULT 25

AAB47447
ID AAB47447 standard; Protein; 353 AA.

XX AAB47447;

XX 31-OCT-2001 (first entry)

XX MOMP P5.

XX surface exposed loop; major outer membrane protein P5; MOMP P5;

XX non-typable H. influenzae; nH1; LBI(f) peptide; B cell epitope;

XX lower respiratory tract infection.

XX Haemophilus influenzae.

XX Key Location/Qualifiers
XX 38..57
XX Domain
XX /label= Loop 1
XX /note= "Extracellular domain"
XX 89..100
XX Domain
XX /label= Loop 2
XX /note= "Extracellular domain"
XX 136..150
XX Domain
XX /label= Loop 3
XX /note= "Extracellular domain"
XX 181..204
XX Domain
XX /label= Loop 4
XX /note= "Extracellular domain"

XX WO200161013-A1.

XX 23-AUG-2001.

XX 13-FEB-2001; 2001MO-EP01556.
 PF
 XX
 PR 15-FEB-2000; 2000GB-0003502.
 XX
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA
 PI Berthel FJ, Denoel P, Poolman J, Thonnard J;
 XX
 XX WPI; 2001-522599/57.
 DR
 XX
 PT Recombinant bacterial outer membrane protein where one or more
 PT surface-exposed loops are modified is useful as a vaccine to prevent or
 PT treat Haemophilus influenzae infection or associated disease, e.g.,
 PT otitis media and conjunctivitis -
 PS
 XX
 PS Disclosure; Fig 1; 29pp; English.
 CC This sequence represents the major outer membrane protein P5 of
 CC non-typable H. influenzae. One or more surface exposed loops of this
 CC protein may be replaced with a modified peptide of the invention. Each
 CC of these peptides contain an LBI(f) peptide which is a 19 amino acid
 CC peptide derived from the sequence of MOMP P5 from strain nH1128,
 CC representing amino acids Arg117 to Gly135. This peptide represents the
 CC third exposed loop of P5 and is a potential B cell epitope. The loops
 CC of the invention are modified in terms of being in a non-native
 CC environment in the recombinant outer membrane protein. The modified
 CC MOMP P5 may be used to induce an immune response in a mammal to
 CC prevent or treat Haemophilus influenzae infection or associated
 CC disease, e.g., otitis media, sinusitis, conjunctivitis, or lower
 CC respiratory tract infection.
 CC
 SQ Sequence 353 AA:
 Query Match 20.5%; Score 176.5; DB 22; Length 353;
 Best Local Similarity 30.6%; Pred. No. 1.2e-10;
 Matches 48; Conservative 27; Mismatches 61; Indels 21; Gaps 4;
 QY 35 PNAPTGY-----TGVIY--TGAPLVNDDETAKLSPSLVYFPDSDSEIKPQA 83
 DB 193 PNTALNTNPMWISINAGISYRGGAAVVAPEVSKTFS-LNSDYTFAGKANKLEKPA 251
 QY 84 AAILDEQAFLTTNOTARLVAGHDEGSRKYNSLGERAAVAVRYLLGKGINQASVE 143
 DB 252 QATLDSIYGEQMSQVSAKVAAGYDRIGSDAFENKLSQERADSVANFEVAKGVAADATIS 311
 QY 144 IISFGERPIAFGTINE-----AMSQNRRLS 171
 DB 312 ATGIGKANFVTGATCDQVNGKRALIACRPDRRVEIA 348
 RESULT 26
 AAR99626
 ID AAR99626 standard; Protein; 214 AA.
 AC AAR99626;
 XX
 XX 01-OCF-1996 (first entry)
 DT
 XX
 DE P. aeruginosa orpF-oprI fusion protein.
 XX
 XX Vaccine; fusion protein; OprF; OprI; outer membrane protein;
 KW antibody; glutathione-S-transferase; GST.
 OS Pseudomonas aeruginosa serotype 6 (ATCC 33354).
 XX
 XX Key Location/Qualifiers
 FT 1..151
 FT Protein /label= oprf
 FT /note= "oprF amino acids 192-342"
 FT 152..214
 FT Protein /label= oprI
 FT /note= "oprI amino acids 21-83"

XX
 PN EP17106-A1.
 XX
 PD 19-JUN-1996.
 XX
 XX 17-NOV-1995; 95EP-0118098.
 PF
 XX 16-DEC-1994; 94EP-0120023.
 PR
 XX (BEHM) BEHRINGERWERKE AG.
 PA
 PI Broeker M, Dondely H, Hungerer K, Knapp B, Ulrich B;
 PI Von Specht BU;
 XX
 XX WPI; 1996-279559/29.
 DR
 DR N-PSDB; AAT32600.
 XX
 XX Fusion proteins for Pseudomonas aeruginosa vaccines - contg.
 PT fragments of outer membrane proteins I and F.
 PS
 XX
 PS Disclosure; Page 13-14; 23pp; English.
 CC A cDNA sequence (AAT32600) codes for a fusion (AAR99626) between amino
 CC acids 192-342 of the outer membrane protein F (oprF) and amino acids
 CC 21-83 of outer membrane protein I (oprI) of Pseudomonas aeruginosa
 CC ATCC 33354. A glutathione-S-transferase fusion with the hybrid
 CC protein was expressed in Escherichia coli and Saccharomyces cerevisiae
 CC transformants. The oprF-oprI hybrid protein, and antibodies raised
 CC against it, conferred protection against P. aeruginosa infection in
 CC laboratory animals. The hybrid protein was significantly more
 CC immunogenic than an oprI-oprF fusion (AAR99627).
 CC
 SQ Sequence 214 AA:
 Query Match 20.4%; Score 175.5; DB 17; Length 214;
 Best Local Similarity 37.5%; Pred. No. 7.6e-11;
 Matches 39; Conservative 16; Mismatches 48; Indels 1; Gaps 1;
 QY 70 YFDDPSDEIKPQAAILDEQAFLTTNOTARLVAGHDEGSRKYNSLGERAAVAVR 129
 DB 51 VKEPDDKSKVENSTADIKNLADPMKQYPSISTVBEHTDSVGTDAVNNKLSRRANAVR 110
 QY 130 NYLLGK-GINQASVELISFGERPIAFGTINEAMSQNRRLS 172
 DB 111 DVLVNEYGVEGGRVNAVGVGSRPVADNATAGRAINRRVSSH 154
 RESULT 27
 AAR85450
 ID AAR85450 standard; Protein; 338 AA.
 AC AAR85450;
 XX
 XX 15-FEB-1996 (first entry)
 DT
 XX
 DE Nontypable H. influenzae P5 protein.
 XX
 XX P5 outer membrane protein; vaccine; otitis media; sinusitis;
 KW chronic pulmonary obstructive disease.
 OS Haemophilus influenzae.
 XX
 XX Key Location/Qualifiers
 FT 15
 FT MISC-difference 195 /note= "amino acid at position 195 is not
 FT /note= "amino acid in the specification"
 FT MISC-difference 311 /note= "amino acid at position 311 is not
 FT /note= "amino acid in the specification"
 FT
 FT EP680765-A1.
 PN
 XX 08-NOV-1995.
 PD

XX 02-MAY-1995; 95EP-0302996.
 PF 05-MAY-1994; 94US-0210394.
 PR (AMCY) AMERICAN CYANAMID CO.
 XX Zlotnick GW;
 XX WPI: 1995-375029/49.
 DR Purified H. influenzae P5 outer membrane protein - used for
 PT preventing reducing susceptibility to or treating H. influenzae
 PT infections
 XX
 PS Disclosure; Page 7-8; 16pp; English.
 CC MontyPable H. influenzae HI outer membrane protein P5 was isolated
 CC by extraction of the outer membrane with detergents and cation-exchange
 CC chromatography. P5 (or its peptide fragments) are used in vaccines
 CC for prevention of H. influenzae infections implicated in otitis media,
 CC sinusitis and chronic pulmonary obstructive disease.
 XX
 SQ Sequence 338 AA;
 Query Match 20.3%; Score 174.5; DB 16; Length 338;
 Best Local Similarity 28.9%; Pred. No. 1.9e-10;
 Matches 46; Conservative 27; Mismatches 57; Indels 29; Gaps 5;
 QY 33 VAPNAPGTG-----TGVIY---TGVAPLVNDDEYKALASLPLSLVYFDPSDEIKP 81
 DB 184 IMPNTAIHNPXISINAGISYRFGGAP-----VKTFSLND--VTFAGKANLKP 234
 QY 82 QAAATLDEQAQFLTNTQARVAVAGHTDERSREYMSLGERRAVAVRNLGKGINQAS 141
 DB 235 QAAATLDSITGEMSGVKSAAKAVAGYDTRIGSDAFNKLQGERADSVANFVAKGVAAAD 234
 QY 142 VEISFGEERPIARETNEAMSQ-----NNRAELS 171
 DB 295 ISATGYGRKANPVTGATXQVGRMALITATLAPDRVEIA 333
 RESULT 28
 AAB59180
 ID AAB59180 standard; protein; 326 AA.
 XX
 AC AAB59180;
 XX
 DT 22-MAR-2001 (first entry)
 XX
 DE Protein associated with C.coli and C.jejuni CadF.
 XX
 KW Fibronectin binding protein; CadF; vaccine; diagnostic assay.
 XX
 OS Pseudomonas fluorescens.
 XX
 FN US6156546-A.
 XX
 PD 05-DEC-2000.
 XX
 PF 15-MAY-1998; 98US-0080025.
 XX
 PR 16-MAY-1997; 97US-0046763.
 XX
 PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
 XX
 PI Garvis SG, Konkel ME;
 XX
 DR WPI: 2001-079546/09.
 XX
 PT Novel isolated polynucleotide useful for producing fibronectin binding
 PT proteins which are useful in production of vaccine, in diagnostic
 PT assays and for prophylactic and therapeutic purposes -

XX Disclosure; Column 39-42; 29pp; English.
 PS
 XX The present invention relates to a Campylobacter jejuni or
 CC Campylobacter coli fibronectin binding protein (CadF). A recombinant
 CC expression vector with cadF is useful in an assay for determining the
 CC presence of C.jejuni or C.coli in a test sample or for
 CC determining whether a test isolate of Campylobacter is a strain
 CC of C.coli. cadF is useful in the construction of DNA probes for
 CC identifying and quantifying the level of expression of CadF in a
 CC cell. The gene can also be used in a vaccine.
 XX
 SQ Sequence 326 AA;
 Query Match 20.2%; Score 174; DB 22; Length 326;
 Best Local Similarity 36.3%; Pred. No. 2e-10;
 Matches 37; Conservative 14; Mismatches 51; Indels 0; Gaps 0;
 QY 70 VFDFPSDEIKPQAAALDEQAQFLTNTQARVAVAGHTDERSREYMSLGERRAVAVR 129
 DB 219 VKFDFKSVKKPNISYGDYKNLADFMAYPATRIVEYAGHTDSIGPDAYNOKLSQRADRVK 278
 QY 130 NYLLGKGINQASVEIISGEEPRPIAFGTNEAMSQNNRAELS 171
 DB 279 QVLVKDGVAPSRITAVGSGSRPVADNATEAGRAVNRREVAS 320
 RESULT 29
 AAR99625
 ID AAR99625 standard; protein; 161 AA.
 XX
 AC AAR99625;
 XX
 DT 01-OCT-1996 (first entry)
 XX
 DE P. aeruginosa oprF C-terminal region (aa190-350).
 XX
 KW Vaccine: fusion protein; OprF; OprI; outer membrane protein;
 KW antibody; glutathione-S-transferase; GST.
 XX
 OS Pseudomonas aeruginosa serotype 6 (ATCC 33554).
 XX
 FN EP717106-A1.
 XX
 PD 19-JUN-1996.
 XX
 PF 17-NOV-1995; 95EP-0118098.
 XX
 PR 16-DEC-1994; 94EP-0120023.
 XX
 PA (BEHM) BEHRINGWERKE AG.
 XX
 PI Broecker M, Domdey H, Hungerer K, Knapp B, Ulrich B;
 PI Von Specht BU;
 XX
 DR WPI: 1996-279559/29.
 XX
 PD N-PSDB; AAT132599.
 XX
 PF Fusion proteins for Pseudomonas aeruginosa vaccines - contg.
 XX fragments of outer membrane proteins I and F.
 XX
 PS Disclosure; Page 11-12; 23pp; English.
 XX
 CC A cdna sequence (AAT132599) codes for the C-terminal region, amino
 CC acids 190-350 (AAR99625), of the Pseudomonas aeruginosa ATCC 33554
 CC outer membrane protein F (oprF). A hybrid protein between
 CC glutathione-S-transferase and the oprF moiety was expressed in
 CC Escherichia coli. OprF-OprI (AAR99626) and OprI-OprF (AAR99627) fusion
 CC proteins were also prep. and tested for their efficacy as vaccines
 CC against P. aeruginosa infection.
 XX
 SQ Sequence 161 AA;

[illegible]

XX		Virulence gene; antibacterial; vaccine; bacterial infection;
KW		septicemia; bronchopneumonia; rhinitis; wound infection.
XX		
CS		Actinobacillus pleuropneumoniae.
XX		
PN		WO20061724-A2.
XX		
PD		19-OCT-2000.
XX		
PF		06-APR-2000; 2000WO-US09218.
XX		
PR		09-APR-1999; 99US-0128689.
XX		10-SEP-1999; 99US-0153453.
PA		(PHAA) PHARMACIA & UPJOHN INC.
PI		
XX		Lowery DE, Fuller TE, Kennedy MJ;
DR		WPI: 2000-647422/62.
XX		N-PSDB; AAC79664.
PT		
PT		Attenuated Pasteurellaceae bacteria comprising mutations in virulence genes, useful as a live attenuated vaccine against bacterial infections
XX		
XX		Claim 39; Pages 308-309; 322pp; English.
CC		The family Pasteurellaceae encompasses several pathogens that infect a wide variety of animals. The present invention relates to virulence genes from Pasteurellaceae. The present sequence is a protein encoded by one such virulence gene. The virulence genes of the present invention may be mutated in order to produce an inactive gene. The inactive virulence gene CC may in turn be used to produce a vaccine, which is useful for treating bacterial infections such as septicemias, bronchopneumonias, rhinitis and wound infections.
SQ		
XX		
Sequence	369 AA:	
Query Match	19.6%;	Score 169; DB 21; Length 369;
Best Local Similarity	35.9%;	Pred. No. 8.5e-10;
Matches	47; Conservative	16; Mismatches 58; Indels 10; Gaps 4.
Oy	34 APNAPTGYTGVIYTF---GVAPLYDNDEIVYKALSKLPISLVYFEDPDSDEIKFOAAALIDE	89
Db	210 AFDHSVTAGLSYRFGGCAVAPVVEEVYTKNFA--FSSDVLDFEKKSLKRAAATLAD-	265
Oy	90 QAOFLITN--OTARVYLVAHGTDERGSGREYNMSLIGERRAVAYNNLLGKINOSEYETIS	146
Db	267 AAMTEIANGLIATPAIAYNVNYTRIGKEASNLKSORRAETVANYLVSRKGOMPANVTAVG	326
Oy	147 FGEERPIAFEGT	157
Db	327 YGEANPYTGAT	337
RESULT 36		
ID	AAR66294	standard; Protein: 359 AA.
AC	AAR66294;	
DT	09-AUG-1995	(first entry)
DE	Non-typable Haemophilus influenza (NTHi) fimbria protein.	
KW	Fimbria protein; vaccine; otitis media.	
OS	Haemophilus influenza strain 1128.	
Key	Location/Qualifiers	
Region	22..33	
FT	/label= amino terminus	

Peptide 234..249
/label= internal CNBR fragment

W09426304-A.

24-NOV-1994.

12-MAY-1994; 94WO-US05477.

18-MAY-1993; 93US-0065442.

(OHIO-) OHIO STATE RES FOUND.

Bakaletz LO, Kolattukudy PE, Sirakova T;

WPI: 1995-006359/01.

N-PSDB: AAQ78916.

Vaccine comprising non-typable Haemophilus influenza fimbria protein - useful in studying, preventing or reducing the severity of otitis media, also fimbria protein and DNA.

Disclosure; Fig 5; 45pp; English.

The fimbria proteins from 15 randomly selected type b and non-typable clinical isolates of Haemophilus influenzae share common epitopes. Thus fimbria isolated from non-typable Haemophilus influenzae 1128 strain is a particularly suitable immunogen to protect against the different non-typable H₂ influenzae that cause otitis media. Fimbria protein is produced by culturing a transformed microbial host, pref. E.coli, Sporodoptera frugiperda or a mucosal pathogen. Fimbria protein (FP) produced by this process is claimed. The FP protein migrates in polyacrylamide gels to a posn. equiv. to a mol. wt. of 25.5 KD or 37.5 KD.

Sequence 359 AA;

Query Match 19.2%; Score 165.5; DB 16; Length 359;
Best Local Similarity 29.9%; Pred. No. 2e-09;
Matches 47; Conservative 26; Mismatches 63; Indels 21; Gaps 4;

QY 35 PNATGY-----TGVY---TGVAPLVNDDETVKALSKLPSTLYFDPDSDEKPPA 83
DB 199 PNTAININPWICGINAGISYRFGQGEAPVVAPEMYSKTFE-LSNDVYTFAFKAMDKPPA 257
QY 84 AAILDEQAQFLITNOTARVIVAGHTDERGSRREYNSLGERRAVAAYKNTLLGKINGQASVE 143
DB 258 QATLDVYGEISQVSKRKAAYAGTITRIGSDAFNKLISOERADSVANFVAKGVAADAIIS 317
QY 144 IISFGERRPIAFGTNEE-----AMSONRRAEIS 171
DB 318 ATGTYGEANPYTGATCDQVKGKRALIACIAPDRRVEIA 354

RESULT 37
AAO17571
ID AAO17571 standard; Protein; 228 AA.

AC AAO17571;

DT 19-JUL-2002 (first entry)

DE M catarrhalis MCA100854 protein SEQ ID NO: 22.

KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;

OS auditory; antibacterial; otitis media; sinusitis; pneumonia.

OS Moraxella catarrhalis.

PN W0200218595-A2.

PD 07-MAR-2002.

PF 28-AUG-2001; 2001WO-CA01221.

PR 28-AUG-2000; 2000US-228294P.

PR 28-AUG-2000; 2000US-228295P.

PR 28-AUG-2000; 2000US-228296P.

PR 29-AUG-2000; 2000US-228438P.

PR 29-AUG-2000; 2000US-228439P.

PR 29-AUG-2000; 2000US-228440P.

PR 29-AUG-2000; 2000US-228441P.

PR 29-AUG-2000; 2000US-228442P.

PR 29-AUG-2000; 2000US-228443P.

PR 29-AUG-2000; 2000US-228511P.

PR 29-AUG-2000; 2000US-228512P.

PR 29-AUG-2000; 2000US-228732P.

PR 01-SEP-2000; 2000US-228773P.

PR 01-SEP-2000; 2000US-229465P.

PR 01-SEP-2000; 2000US-229474P.

PR 01-SEP-2000; 2000US-229475P.

PR 05-SEP-2000; 2000US-229478P.

PR 05-SEP-2000; 2000US-229740P.

PR 05-SEP-2000; 2000US-229803P.

PR 05-SEP-2000; 2000US-229805P.

PR 05-SEP-2000; 2000US-229806P.

PR 05-SEP-2000; 2000US-229809P.

PR 06-SEP-2000; 2000US-229811P.

PR 06-SEP-2000; 2000US-230214P.

PR 06-SEP-2000; 2000US-230250P.

PR 06-SEP-2000; 2000US-230252P.

PA (AVET) AVENTIS PASTEUR LTD.

PI Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;

DR WPI: 2002-401721/43.

XX N-PSDB: AAL46503.

XX Moraxella polypeptide and polynucleotides useful as vaccine for immunizing a host e.g. humans against disease e.g. otitis media, pneumonia, caused by infection of the bacteria

XX Claim 28; Fig 21; 277pp; English.

The present invention provides the protein and coding sequences of proteins from Moraxella catarrhalis. These can be used to produce vaccines which protect against M. catarrhalis infection, which can cause otitis media, respiratory infection, sinusitis, and pneumonia. The present sequence is a protein of the invention.

Sequence 228 AA;

Query Match 19.2%; Score 165; DB 23; Length 228;
Best Local Similarity 30.7%; Pred. No. 1.2e-09;
Matches 42; Conservative 22; Mismatches 71; Indels 2; Gaps 1;

QY 33 VAPNAPFTGYGVYGVAPLVNDDETVKALSKLPSTLYFDPDSDEKPPAAILDEQAQ 92
DB 68 VADSIETARVAIVAGC--DYEENMDLILNALTQIINFALDSTETPOENKKEIIDAAE 125

QY 93 FLITNOTARVIVAGHTDERGSRREYNSLGERRAVAAYKNTLLGKINGQASVEIISFGGERP 152
DB 126 KIKAVPEETTLRIHTDQGTHEYNOPLSESRRAAAYVEIYLSKVAERLNTOGASPDY 185

QY 153 IAFGTNEAMSONRRAE 169
DB 186 VASNATEGGRFQNRRIE 202

RESULT 38

AAAS0718
ID AAAS0718 standard; Protein; 344 AA.

AC AAAS0718;

XX 18-APR-2002 (first entry)
 XX Haemophilus paragallinarum strain 2403 haemagglutinin.
 XX DE Haemagglutinin; Hagg; antigen; vaccine; immunisation; coryza;
 XX KW chicken.
 XX OS Haemophilus paragallinarum.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /Label= Signal_peptide
 FT 22..344
 FT Protein /Label= Mature_protein
 XX
 PN WO200204485-A1.
 PD 17-JAN-2002.
 XX
 PD 06-JUL-2001; 2001WO-AU00822.
 XX
 PF 07-JUL-2000; 2000AU-0008652.
 XX
 PR (UYOU) UNIV QUEENSLAND.
 XX
 PA Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
 PI WPI; 2002-154917/20.
 XX
 DR N-PSDB; ABA91419.
 XX
 DR New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
 PT useful as a vaccine for immunising chickens against coryza caused by
 PT the species -
 PS
 PS Claim 2; Fig 4; 67pp; English.
 XX
 CC The present sequence is that of the Hagg haemagglutinin of
 CC Haemophilus paragallinarum strain 2403 (serovar A), the causative
 CC agent of infectious coryza in chickens. The invention provides
 CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
 CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
 CC A, B and C) of H. paragallinarum, the causative agent of infectious
 CC coryza of chickens. The polypeptides are useful in vaccines for
 CC immunisation against infectious coryza, as are the nucleic acids
 CC when expressed in attenuated bacteria, especially Salmonella or
 CC Mycoplasma (claimed). The recombinant polypeptide is preferably
 CC the mature protein, or a biologically active fragment, variant or
 CC derivative, that is capable of eliciting an immune response,
 CC providing protection against one or more strains of H.
 CC paragallinarum in chickens. Also claimed are methods of using the
 CC haemagglutinin polypeptides and nucleic acids for detection and
 CC diagnosis of infectious coryza in chickens.
 CC
 SO Sequence 344 AA:
 Query Match 18.8%; Score 161.5; DB 23; Length 344;
 Best Local Similarity 30.9%; Pred. No. 5.1e-09;
 Matches 51; Conservative 25; Mismatches 68; Indels 21; Gaps 5;
 QY 22 GCANKSTQVAVAPNAPTYGTYIT--GVAPLYDNDNETYKALASKLPISLVFDFSDDEI 79
 DB 181 GYRKDGSRVDYTPSIGVTAGLSYRFQSAPEVPEKVVAKTFA--LNSDVTFAFGKANL 238
 QY 80 KFOAAALID-----EQAQFLTTNQTARVLVAGHTDEGSRREYNMSLGERRAVAVRNYLGG 135
 DB 239 RPEAQNVLDTGIGYETIOL-----KSYOVVAGTYTDIGSEANLKSORRADTVANYLYSK 294
 QY 136 GINQASVEIISFGEERPIAFGTFNE-----EAMSONRRAEIS 171
 DB 295 GVAQEVISSTGYGEANPVGACDVPVKKRALLIACLAADRYREIS 339

RESULT 39
 AAM50721
 ID AAM50721 standard; Protein; 344 AA.
 XX
 AC AAM50721;
 XX
 DT 18-APR-2002 (first entry)
 XX
 DE Haemophilus paragallinarum strain 0222 haemagglutinin.
 XX
 KW Haemagglutinin; Hagg; antigen; vaccine; immunisation; coryza;
 XX chicken.
 XX OS Haemophilus paragallinarum.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /Label= Signal_peptide
 FT 22..344
 FT Protein /Label= Mature_protein
 XX
 PN WO200204485-A1.
 PD 17-JAN-2002.
 XX
 PD 06-JUL-2001; 2001WO-AU00822.
 XX
 PF 07-JUL-2000; 2000AU-0008652.
 XX
 PR (UYOU) UNIV QUEENSLAND.
 XX
 PA Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
 PI WPI; 2002-154917/20.
 XX
 DR N-PSDB; ABA91422.
 XX
 DR New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
 PT useful as a vaccine for immunising chickens against coryza caused by
 PT the species -
 PS
 PS Claim 2; Fig 4; 67pp; English.
 XX
 CC The present sequence is that of the Hagg haemagglutinin of
 CC Haemophilus paragallinarum strain 0222 (serovar B), the causative
 CC agent of infectious coryza in chickens. The invention provides
 CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
 CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
 CC A, B and C) of H. paragallinarum, the causative agent of infectious
 CC coryza of chickens. The polypeptides are useful in vaccines for
 CC immunisation against infectious coryza, as are the nucleic acids
 CC when expressed in attenuated bacteria, especially Salmonella or
 CC Mycoplasma (claimed). The recombinant polypeptide is preferably
 CC the mature protein, or a biologically active fragment, variant or
 CC derivative, that is capable of eliciting an immune response,
 CC providing protection against one or more strains of H.
 CC paragallinarum in chickens. Also claimed are methods of using the
 CC haemagglutinin polypeptides and nucleic acids for detection and
 CC diagnosis of infectious coryza in chickens.
 CC
 SO Sequence 344 AA:
 Query Match 18.8%; Score 161.5; DB 23; Length 344;
 Best Local Similarity 30.9%; Pred. No. 5.1e-09;
 Matches 51; Conservative 25; Mismatches 68; Indels 21; Gaps 5;
 QY 22 GCANKSTQVAVAPNAPTYGTYIT--GVAPLYDNDNETYKALASKLPISLVFDFSDDEI 79
 DB 181 GYRKDGSRVDYTPSIGVTAGLSYRFQSAPEVPEKVVAKTFA--LNSDVTFAFGKANL 238
 QY 80 KFOAAALID-----EQAQFLTTNQTARVLVAGHTDEGSRREYNMSLGERRAVAVRNYLGG 135
 DB 239 RPEAQNVLDTGIGYETIOL-----KSYOVVAGTYTDIGSEANLKSORRADTVANYLYSK 294

QY 136 GINQASVEIISFGEEPIAFGTNE-----EAMSONRRAELS 171
 DB 295 GVAQEVISSTGYGEANPVTGAKCDYTKGRKALACIADDRREIS 339

RESULT 40

AAM50722
 ID AAM50722 standard; Protein; 344 AA.
 AC AAM50722;
 DT 18-APR-2002 (first entry)

DE Haemophilus paragallinarum strain 2671 haemagglutinin.
 XX
 KW Haemagglutinin; Hagg; antigen; vaccine; immunisation; coryza;
 XX chicken.

OS Haemophilus paragallinarum.

FH Key Location/Qualifiers

FT Peptide 1..21

FT /label= Signal_peptide

FT Protein 22..344

PN /label= Mature_protein

PD WO200204485-A1.

PF 17-JAN-2002.

PR 06-JUL-2001; 2001WO-AU00822.

PA 07-JUL-2000; 2000AU-0008652.

PI (UYOU) UNIV QUEENSLAND.

DR Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;

DR WPI; 2002-154917/20.

PT N-PSDB; ABA91423.

PS New haemagglutinin polypeptide isolated from Haemophilus paragallinarum

XX the species - useful as a vaccine for immunising chickens against coryza caused by

XX Claim 2; Fig 4; 67pp; English.

CC The present sequence is that of the Hagg haemagglutinin of

CC Haemophilus paragallinarum strain 2671 (serovar B), the causative

CC agent of infectious coryza in chickens. The invention provides

CC recombinant haemagglutinin polypeptides (see AAM50716-27) and

CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars

CC A, B and C) of H. paragallinarum, the causative agent of infectious

CC coryza of chickens. The polypeptides are useful in vaccines for

CC immunisation against infectious coryza, as are the nucleic acids

CC Mycoplasma (claimed). The recombinant polypeptide is preferably

CC the mature protein, or a biologically active fragment, variant or

CC derivative, that is capable of eliciting an immune response.

CC providing protection against one or more strains of H.

CC paragallinarum in chickens. Also claimed are methods of using the

CC haemagglutinin polypeptides and nucleic acids for detection and

CC diagnosis of infectious coryza in chickens.

XX Sequence 344 AA;

QY Query Match

Best Local Similarity 18.8%; Score 161.5; DB 23; Length 344;

Matches 51; Conservative 25; Mismatches 68; Indels 21; Gaps 5;

DB 22 GCANKSTQVAVANAPTGTGTGYIT--GVAAPLVNDDEYKALASKIPSLVYFDPDSDEI 79

181 GRVKGDSRVDTYPTISGTVAGLSYRGGAPVPEKPVAKTFA--LNSDVTFAFGKANL 238

QY 80 KPOAAALD-----EQAQFLTTNOTARVLVAGHTDGRSGREYNMSIGERRAVAVRNYLLGK 135
 DB 239 RPEAQNVLDGIYGEIQAQL-----KSYQVDVAGYTDRIISGEANLKLISORRADTAVANYLVSK 294
 QY 136 GINQASVEIISFGEEPIAFGTNE-----EAMSONRRAELS 171
 DB 295 GVAQEVISSTGYGEANPVTGAKCDYTKGRKALACIADDRREIS 339

Search completed: July 6, 2003, 14:04:28
 Job time : 64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:14:15 ; Search time 16 Seconds

(without alignments)
445.871 Million cell updates/sec

Title: US-09-674-779B-2

Perfect score: 172

Sequence: 1 MMLHIQIAAAALSVLTFM.....IAFGTNEAWSONRRRLSLSY 172

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Word size : 0

Search: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	7.6	176	1	PAL_LEGN
2	8	4.7	491	1	PBP_BACSU
3	8	4.7	829	1	E74A_DROME
4	7	4.1	37	1	ANP3_PSEAM
5	7	4.1	40	1	ANP8_MYOAE
6	7	4.1	82	1	ANPA_PSEAM
7	7	4.1	120	1	CU19_ARADI
8	7	4.1	127	1	CU26_ARADI
9	7	4.1	145	1	AZUP_PARP
10	7	4.1	173	1	PAL_ECOLI
11	7	4.1	196	1	RBMS_XENLA
12	7	4.1	200	1	RBMS_CHICK
13	7	4.1	203	1	CPPL_MOUSE
14	7	4.1	213	1	RL14_RAT
15	7	4.1	251	1	COB1_MYCTU
16	7	4.1	261	1	GSHI_MOUSE
17	7	4.1	262	1	MYRK_METRA
18	7	4.1	266	1	MTRK_METTH
19	7	4.1	266	1	MTRK_METTH
20	7	4.1	290	1	HXB8_HUMAN
21	7	4.1	291	1	PSBO_CHLRE
22	7	4.1	294	1	GBSK_BACHD
23	7	4.1	304	1	GSH2_HUMAN
24	7	4.1	305	1	GSH2_MOUSE
25	7	4.1	322	1	GVPL_HAIME
26	7	4.1	322	1	SLPL_DROME
27	7	4.1	325	1	AB5B_MYCKA
28	7	4.1	330	1	GDC_BOVIN
29	7	4.1	338	1	RLAO_METJA
30	7	4.1	349	1	GUB_FIBSU
31	7	4.1	350	1	DKK3_HUMAN
32	7	4.1	354	1	KLF2_MOUSE
33	7	4.1	355	1	KLF2_HUMAN

RESULT 1

ALIGNMENTS

34	7	4.1	367	1	BET3_MESAU	009029 mesocricetu
35	7	4.1	370	1	YCBU_BACSU	P42253 bacillus su
36	7	4.1	389	1	FL_ORYSA	O24175 oryza sativ
37	7	4.1	401	1	HB9_HUMAN	P50219 homo sapien
38	7	4.1	409	1	NU4M_ASCSU	P24880 ascaris suu
39	7	4.1	427	1	RL4_HUMAN	P36578 homo sapien
40	7	4.1	438	1	TUD4_AGRVI	O44472 agrobacteri
41	7	4.1	443	1	MURE_STRO	O33804 streptomyce
42	7	4.1	475	1	EYX2_MOUSE	P49749 mus musculu
43	7	4.1	476	1	EYX2_HUMAN	O03828 homo sapien
44	7	4.1	478	1	FXD3_HUMAN	O94155 homo sapien
45	7	4.1	481	1	KNIR_DROYI	O24753 drosophila
46	7	4.1	481	1	BHB3_HUMAN	O95019 homo sapien
47	7	4.1	485	1	NAB1_MESAU	O95058 mesocricetu
48	7	4.1	485	1	NAB1_RAT	O62722 rattus norv
49	7	4.1	486	1	NAB1_HUMAN	O13506 homo sapien
50	7	4.1	486	1	NAB1_MOUSE	O61122 mus musculu
51	7	4.1	492	1	TI5D_HUMAN	P47974 homo sapien
52	7	4.1	496	1	BAFL_KLUMA	P33293 kluyveromyc
53	7	4.1	498	1	CAUR_PIG	P25117 sus scrofa
54	7	4.1	498	1	LYXK_ECOLI	P37677 escherichia
55	7	4.1	503	1	TGR1_HUMAN	P36897 homo sapien
56	7	4.1	517	1	VSM5_TRYBR	P21840 trypanosoma
57	7	4.1	530	1	ZIC2_MOUSE	O62520 mus musculu
58	7	4.1	532	1	ZIC2_HUMAN	O95409 homo sapien
59	7	4.1	537	1	MMSA_BOVIN	O07536 bos taurus
60	7	4.1	547	1	MCPC_SALTY	O02755 salmonella
61	7	4.1	551	1	MCPI_ECOLI	P02942 escherichia
62	7	4.1	553	1	MCPI_ECOLI	P07017 escherichia
63	7	4.1	553	1	MCPI_ECOLI	P02941 salmonella
64	7	4.1	622	1	SR68_CANFA	O00004 canis fam11
65	7	4.1	647	1	KNRL_DROME	P13054 drosophila
66	7	4.1	665	1	LI23_CAEEL	O00990 caenorhabdi
67	7	4.1	671	1	HMOC_DROME	P22810 drosophila
68	7	4.1	688	1	YHTO_YEAST	P38708 saccharomyc
69	7	4.1	741	1	ERF2_PICPI	P23637 pichia pinu
70	7	4.1	766	1	EVA_DROME	O05201 drosophila
71	7	4.1	810	1	REB1_YEAST	P21538 saccharomyc
72	7	4.1	883	1	E74B_DROME	P11536 drosophila
73	7	4.1	884	1	YAVK_CAEEL	P17851 caenorhabdi
74	7	4.1	912	1	KPCM_HUMAN	O15139 homo sapien
75	7	4.1	918	1	KPCM_MOUSE	O62101 mus musculu
76	7	4.1	988	1	OMB_DROME	O24432 drosophila
77	7	4.1	1047	1	CUGA_ECOLI	P38054 escherichia
78	7	4.1	1092	1	DP2L_METTH	O27579 methanobact
79	7	4.1	1225	1	CTD2_HUMAN	O94ub3 homo sapien
80	7	4.1	1385	1	NAC2_CHLRE	O91ame8 chlamydomon
81	7	4.1	2038	1	FSH_DROME	P13709 drosophila
82	7	4.1	2175	1	HMCU_DROME	P10180 drosophila
83	7	4.1	2799	1	EDD_HUMAN	O95071 homo sapien
84	7	3.5	33	1	ANP3_MYOSC	O04367 myoxocephal
85	6	3.5	45	1	ANP8_MYOAE	P04368 myoxocephal
86	6	3.5	45	1	ANP8_MYOAE	P04368 myoxocephal
87	6	3.5	52	1	RL4O_YEAST	P14796 saccharomyc
88	6	3.5	53	1	RL4O_CRYNE	P40909 cryptococcu
89	6	3.5	55	1	RL32_VIBCH	O94k37 vibrio chol
90	6	3.5	72	1	OMP_LODMI	P80045 locusta mig
91	6	3.5	75	1	RL4O_TOBAC	P19379 nicotiana t
92	6	3.5	74	1	SR14_MACRA	O18881 macera radi
93	6	3.5	85	1	ANP4_PSEAM	P02734 pseudopleur
94	6	3.5	85	1	MRKD_RAT	P20468 rattus norv
95	6	3.5	87	1	YHCN_ECOLI	P46477 escherichia
96	6	3.5	90	1	DBHA_AERHY	O91966 aeromonas h
97	6	3.5	90	1	DBHA_VIBPR	P20800 vibrio prot
98	6	3.5	91	1	ANPX_PSEAM	P07835 pseudopleur
99	6	3.5	91	1	ANPX_PSEAM	P23899 pseudopleur
100	6	3.5	97	1	ANP_LTIME	P09031 limanda fer

PAL_LEGPN
ID PAL_LEGPN STANDARD; PRT; 176 AA.
AC P26493;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Peptidoglycan-associated lipoprotein precursor (19 kDa surface antigen) (PPL).
GN PAL OR PPLA.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A100 / Serogroup 1;
RC MEDLINE=92114778; PubMed=1766377;
RA Engleberg N.C., Howe D.C., Rogers J.E., Arroyo J., Eisenstein B.I.;
RT "Characterization of a Legionella pneumophila gene encoding a
RT lipoprotein antigen."
RL MOL. Microbiol. 5:2021-2029(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=91310296; PubMed=1855972;
RA Ludwig B., Schmid A., Maire R., Hacker J.;
RT "Cloning, genetic analysis, and nucleotide sequence of a determinant
RT coding for a 19-kilodalton peptidoglycan-associated protein (Ppl) of
RT Legionella pneumophila."
RL Infect. Immun. 59:2515-2521(1991).
CC -1- FUNCTION: VERY STRONGLY ASSOCIATED WITH THE PEPTIDOGLYCAN.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -1- SIMILARITY: TO OTHER PAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X60543; CAA43033.1; -
DR PIR: A60337; A60337.
DR PIR: S16631; S16631.
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_OmpA; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS01068; OMPA; 1.
KW Lipoprotein; Outer membrane; Antigen; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 176 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
FT LIPID 22 22 N-ACYL DIGLYCERIDE (PROBABLE).
FT DOMAIN 104 148 OMPA-LIKE.
SQ SEQUENCE 176 AA; 18911 MW; 7D9C3BEECBEE21DB CRC64;
Query Match 7.6%; Score 13; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 5,2e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative penicillin binding protein precursor.
GN PBP OR DACC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168.
RC MEDLINE=95227362; PubMed=7711903;
RA Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,
RA Grandi G.;
RT "A putative new peptide synthase operon in Bacillus subtilis: partial
RT characterization."
RL Microbiology 141:645-648(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berrtero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschel C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klauer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kunano M.,
RA Kurita K., Lapins A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Meliado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Prescan E., Puig C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Seldguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccon E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Welleneger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S13; ALSO KNOWN AS THE
CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 3 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z34883; CAA84366.1; -
DR EMBL: Z99113; CAB13718.1; -
DR MEROPS: S13.0PW; -
DR Subtilist: BG10969; pdp.
DR InterPro: IPR000667; Peptidase_S13.
DR Pfam: PF02113; Peptidase_S13; 1.
DR PRINTS: PR00922; DADACBPASE3.
DR TIGRFAMS: TIGR00666; pBP4; 1.
KW Peptidoglycan synthesis; Cell division; Cell wall; Hydrolase; Signal;
KW Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 491 PUTATIVE PENICILLIN BINDING PROTEIN.

Db 6 AAAAAAL 12

RESULT 5

ANP8_MYOAE STANDARD: PRT: 40 AA.
 AC P20617;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 DE Antifreeze peptide GS-8.
 OS Myoxocephalus aeneus (Grubby sculpin).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
 OC Cottidae; Cottidae; Myoxocephalus.
 OC NCBI_TaxID=8096;
 RN [1]
 RP Chakrabarty A., Hew C.-L., Shears M., Fletcher G.;
 RT "Primary structures of the alanine-rich antifreeze polypeptides from
 RT Grubby sculpin, Myoxocephalus aeneus";
 RL Can. J. Zool. 66:403-408(1988).
 CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE I AFP ARE
 CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
 CC PIR: S07046; FDF18G.
 DR Antifreeze protein; Repeat. BLOCKED.
 KW MOD_RES 1
 FT MOD_RES 1
 SQ SEQUENCE 40 AA; 3579 MM; 32F50E243C2AD11 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 40;
 Best Local Similarity 100.0%; Pged. No. 6; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

OY 8 AAAAAAL 14
 Db 14 AAAAAAL 20

RESULT 6

ANP8_PSEAM STANDARD: PRT: 82 AA.
 AC P04002;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antifreeze protein A/B precursor.
 OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes
 americanus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 OC Pleuronectidae; Pleuronectidae; Pseudopleuronectes.
 OC NCBI_TaxID=8265;
 RN [1]
 RP SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=82197490; PubMed=6952188;
 RA Davies P.L., Roach A.H., Hew C.-L.;
 RT "DNA sequence coding for an antifreeze protein precursor from winter
 RT flounder";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).
 RN [2]
 RP SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=8259236; PubMed=3133486;
 RA Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
 RT "Differential amplification of antifreeze protein genes in the
 RT pleuronectidae";
 RL J. Mol. Evol. 27:29-35(1988).
 RN [3]
 RP SEQUENCE FROM N.A. (PROTEIN B).
 RX MEDLINE=84264559; PubMed=6086629;

RA Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
 RT "Antifreeze protein genes of the winter flounder";
 RL J. Biol. Chem. 259:9241-9247(1984).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92209995; PubMed=1555765;
 RA Davies P.L.;
 RT "Conservation of antifreeze protein-encoding genes in tandem
 RT repeats";
 RL Gene 112:163-170(1992).
 RN [5]
 RP 3D-STRUCTURE MODELING OF 45-81.
 RX MEDLINE=92148833; PubMed=1738160;
 RA Chou K.-C.;
 RT "Energy-optimized structure of antifreeze protein and its binding
 RT mechanism";
 RL J. Mol. Biol. 223:509-517(1992).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
 RX MEDLINE=95281060; PubMed=7760940;
 RA Sticherl F., Yang D.S.C.;
 RT "Ice-binding structure and mechanism of an antifreeze protein from
 RT winter flounder";
 RL Nature 375:427-431(1995).

CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PROTEIN A.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE I AFP ARE
 CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: L00138; AAB59964.1;
 CC EMBL: L29178; AAB59964.1; JOINED.
 CC EMBL: M62414; AAA49469.1;
 CC EMBL: X07506; CAA30389.1;
 CC EMBL: M62416; AAA49471.1;
 CC EMBL: M62417; AAA49472.1;
 CC PIR: A03194; FDF1AW.
 CC PIR: A03161; A05161.
 CC PIR: S02326; S02326.
 CC PIR: JS0704; JS0704.
 CC PDB: 1ATF; 15-OCT-94.
 CC PDB: 1WEA; 03-JUN-95.
 CC PDB: 1WEA; 03-JUN-95.
 CC InterPro: IPR000104; Antifreeze_1.
 CC PRINTS: PR00308; ANTIFREEZE1.
 KW Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
 FT SIGNAL 1
 FT PROPEP 22
 FT CHAIN 45
 FT VARIANT 36
 FT VARIANT 70
 FT CONFLICT 24
 FT HELIX 46
 FT SEQUENCE 82 AA; 7711 MM; C2AE7B74C0D46CC1 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 82;
 Best Local Similarity 100.0%; Pged. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7

OY 8 AAAAAAL 14
 Db 50 AAAAAAL 56

CUI9_ARADI
ID CUI9_ARADI STANDARD; PRT; 120 AA.
AC P80515;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adult-specific rigid cuticular protein 11.9 (ACP 11.9).
OS Araneus diadematus (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.
NCBI_TaxID=45920;
RN [1]
RP SEQUENCE.
RC TISSUE=Cuticle;
RX MEDLINE=9716616; PubMed=9014336;
RA Norup T., Berg T., Stenholm H., Andersen S.O., Hoerjup P.;
RT Purification and characterization of five cuticular proteins from
the spider Araneus diadematus.;
RL Insect Biochem. Mol. Biol. 26:907-915(1996).
CC -1- FUNCTION: COMPONENT OF THE RIGID CUTICLE OF THE SPIDER.
CC -1- MASS SPECTROMETRY: MW=11872.1; METHOD=Electrospray.
DR InterPro: IPR000618; Insect_cuticle.
DR Pfam: PF00379; Chitin_bind_4; 1.
DR PRINTS: PR00947; CUTICLE.
DR PROSITE: PS00233; CUTICLE; 1.
KM Structural protein; Cuticle.
FT DOMAIN 73
SQ SEQUENCE 120 AA; 11871 MW; 803E577C61999DC2 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
Db 73 AAAAAA 79

RESULT 8
CUI9_ARADI
ID CUI9_ARADI STANDARD; PRT; 127 AA.
AC P80517;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adult-specific rigid cuticular protein 12.6 (ACP 12.6).
OS Araneus diadematus (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.
NCBI_TaxID=45920;
RN [1]
RP SEQUENCE.
RC TISSUE=Cuticle;
RX MEDLINE=9716616; PubMed=9014336;
RA Norup T., Berg T., Stenholm H., Andersen S.O., Hoerjup P.;
RT Purification and characterization of five cuticular proteins from
the spider Araneus diadematus.;
RL Insect Biochem. Mol. Biol. 26:907-915(1996).
CC -1- FUNCTION: COMPONENT OF THE RIGID CUTICLE OF THE SPIDER.
CC -1- MASS SPECTROMETRY: MW=12587.5; METHOD=Electrospray.
DR InterPro: IPR000618; Insect_cuticle.
DR Pfam: PF00379; Chitin_bind_4; 1.
DR PRINTS: PR00947; CUTICLE.
DR PROSITE: PS00233; CUTICLE; 1.
KM Structural protein; Cuticle.
FT DOMAIN 73
SQ SEQUENCE 127 AA; 12589 MW; 9BFBCEB2B6AAB9 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
Db 73 AAAAAA 79

RESULT 9
AZUP_PARP_N
ID AZUP_PARP_N STANDARD; PRT; 145 AA.
AC P80401;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pseudocazurin precursor.
OS PADS.
OS Paracoccus pantotrophus (Rhizosphere pantotropha).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group.
OC Paracoccus.
NCBI_TaxID=82367;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-32.
RC STRAIN=ATCC 35512 / LMD 82.5;
RX MEDLINE=97184655; PubMed=9032456;
RA Leung Y.-C., Chan C., Reader J.S., Willis A.C., van Spanning R.J.M.,
RA Ferguson S.J., Radford S.E.,
RT "The pseudocazurin gene from Rhizosphere pantotropha: analysis of
upstream putative regulatory sequences and overexpression in
Escherichia coli.";
RL Biochem. J. 321:699-705(1997).
RN [2]
RP SEQUENCE OF 23-145.
RC STRAIN=ATCC 35512 / LMD 82.5;
RX MEDLINE=95289994; PubMed=7772045;
RA Chan C., Willis A.C., Robinson C.V., Aplin R.T., Radford S.E.,
RA Ferguson S.J.;
RT "The complete amino acid sequence confirms the presence of
pseudocazurin in Rhizosphere pantotropha.";
RL Biochem. J. 308:585-590(1995).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RA Williams P.A.;
RL Submitted (FEB-1997) to the PDB data bank.
CC -1- FUNCTION: THIS SOLUBLE ELECTRON TRANSFER COPPER PROTEIN IS
REQUIRED FOR THE INACTIVATION OF COPPER-CONTAINING NITRITE
REDUCTASE IN THE PRESENCE OF OXYGEN.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
EMBL: Z73141; CA97485.1; -
EMBL: Z70033; CA93848.1; -
DR PDB: 1ADM; 15-MAY-97.
DR InterPro: IPR000923; BlueCu_1.
DR InterPro: IPR001235; Copper_blue.
DR Pfam: PF00127; copper_bind; 1.
DR PRINTS: PR00156; COPPERBLUE.
DR ProDom: PD001235; COPPERBLUE.
DR PROSITE: PS00196; COPPER_BLUE; 1.
DR Copper; Electron transport; Signal; Periplasmic; 3D-structure.
FT SIGNAL 1 22
FT CHAIN 23 145 PSEUDOCAZURIN.
FT DOMAIN 27 115 PLASTOCYANIN-LIKE.
FT METAL 62 62 COPPER (BY SIMILARITY).
FT METAL 100 100 COPPER (BY SIMILARITY).
FT METAL 103 103 COPPER (BY SIMILARITY).

FT METAL 108 108 COPPER (BY SIMILARITY).
 SQ SEQUENCE 145 AA; 15446 MW; 725ECB5929EC3831 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 145;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAL 14
 111111
 7 AAAAAAL 13

RESULT 10
 ID PAL_ECOLI STANDARD; PRT; 173 AA.
 AC P07176;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL OR EXCC OR B0741 OR Z0909 OR ECS0776.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=87133578; PubMed=3545827;
 RA Chen R., Henning U.;
 RT "Nucleotide sequence of the gene for the peptidoglycan-associated
 RT lipoprotein of Escherichia coli K12.";
 RL Eur. J. Biochem. 163:73-77(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=92244043; PubMed=1574003;
 RA Lazzeroni J.-C., Portallier R.;
 RT "The excc gene of Escherichia coli K-12 required for cell envelope
 RT integrity encodes the peptidoglycan-associated lipoprotein (Pnl).";
 RL Mol. Microbiol. 6:735-742(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alta H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horinouchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 RA Iida T., Takami H., Honda T., Hattori M., Shinozawa H.,
 RA Kuwara S., Shiba T., Hattori M., Shinozawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RP SEQUENCE OF 1-30 FROM N.A.
 RC STRAIN-JM105;
 RX MEDLINE=90078104; PubMed=2687247;
 RA Levengood S.K., Webster R.E.;
 RT "Nucleotide sequences of the tolA and tolB genes and localization of
 RT their products, components of a multistep translocation system in
 RT Escherichia coli.";
 RL J. Bacteriol. 171:6600-6609(1989).
 RN [8]
 RP CRYSTALLIZATION.
 RX MEDLINE=21111468; PubMed=11173492;
 RA Abargel C., Walburger A., Chenivesse S., Lazdunski C.;
 RT "Crystallization and preliminary crystallographic study of the
 RT peptidoglycan-associated lipoprotein from Escherichia coli.";
 RL Acta Crystallogr. D 57:117-119(2001).
 CC -1- FUNCTION: Thought to play a role in bacterial envelope integrity.
 CC -1- Very strongly associated with the peptidoglycan.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor.
 CC -1- SIMILARITY: TO OTHER PNL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X05123; CAA28771.1; -;
 CC EMBL: X65796; CAA46773.1; -;
 CC EMBL: AE000177; AAC73835.1; -;
 CC EMBL: D90713; BAA35407.1; -;
 CC EMBL: AE005252; AAG55077.1; -;
 CC EMBL: AP002553; BAB34199.1; -;
 CC EMBL: M28232; -, NOT_ANNOTATED_CDS.
 CC PIR: A27534; LPECPG.
 CC PIR: S20547; S20547.
 CC ECGene: EG10684; pal.
 CC InterPro: IPR001145; Bac_OmpA.
 CC Pfam: PF00691; OmpA; 1.
 CC PRINTS: PR01021; OMPADOMAIN.
 CC PRODOM: PD000930; Bac_OmpA; 1.
 CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC PROSITE: PS01068; OMPA; 1.
 CC Outer membrane; Signal; Lipoprotein; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 173 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
 FT LIPID 22 22 N-ACYL DIGLYCERIDE.
 FT DOMAIN 105 149 OMPA-LIKE.
 FT SEQUENCE 173 AA; 18824 MW; 449F9595C0274430 CRC64;
 SQ SEQUENCE 173 AA; 18824 MW; 449F9595C0274430 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 173;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SLGERRA 125

Db 120 SLGERRA 126

RESULT 11

RBMS_XENLA STANDARD; PRT; 196 AA.

AC Q9YGP5;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE RNA-binding protein with multiple splicing homolog (RBP-MS) (Heart,

DE RRM Expressed Sequence) (Hermes).

GN RBPM5 OR HERMES.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

OC Xenopodidae; Xenopus.

OX NCBI_TaxID=8335;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=99195799; PubMed=1096065;

RA Gerber W.V., Yatskevych T.A., Antin P.B., Correia K.M., Conlon R.A.,

RT Kriegl P.A.;

RT "The RNA-binding protein gene, hermes, is expressed at high levels in

RT the developing heart."

RL Mech. Dev. 80:77-86(1999).

CC -1- FUNCTION: MAY BIND RNA.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING HEART, PRONEPHROS,

CC RETINA AND EPIPHYSIS. IN ADULT, HIGH EXPRESSION IN HEART, MODERATE

CC IN KIDNEY, UNDETECTABLE IN LIVER, LONG AND SKELETAL MUSCLE.

CC -1- DEVELOPMENTAL STAGE: MRNA FIRST DETECTED IN THE TAILBUD EMBRYO

CC (STAGE 26) IN THE PAIRED HEART PRIMORDIA AND IN THE CONDENSING

CC EPITHELIUM THAT WILL FORM THE PRONEPHROS; AT THE LATE TAILBUD

CC STAGE (STAGE 34) IN THE DEVELOPING RETINA AND EPIPHYSIS. AS

CC DEVELOPMENT PROCEEDS, DETECTED THROUGH THE ENTIRE LENGTH OF THE

CC HEART TUBE, IN THE MUSCULAR TISSUE OF THE OUTFLOW TRACT, AND IN

CC THE DUCT EPITHELIUM OF THE PRONEPHROS. DURING LATER DEVELOPMENT,

CC MRNA FOUND IN ALL SUBREGIONS OF THE HEART, IN THE GLOMUS, TUBULES

CC AND DUCT OF THE PRONEPHROS, IN THE RETINAL GANGLION CELL LAYER

CC (GCL) AND IN THE EPIPHYSIS.

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF107889; AAD16971.1; -

DR HSSP: P09012; 201A.

DR InterPro: IPR000504; RNA_rec_mot.

DR Pfam: PF00076; rrm; 1.

DR SMART: SM00360; RRM; 1.

DR PROSITE: PS50102; RRM; 1.

DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.

RN RNA-binding.

FT DOMAIN 20 97 RNA-BINDING (RRM).

FT DOMAIN 11 14 POLY-ASN.

FT DOMAIN 167 172 POLY-ALA.

SQ SEQUENCE 196 AA; 21675 MW; 7964198C1122A3D3 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 196;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 AAAAAL 14

Db 167 AAAAAL 173

RESULT 12

RBMS_CHICK STANDARD; PRT; 200 AA.

AC Q9W611;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE RNA-binding protein with multiple splicing homolog (RBP-MS) (Heart,

DE RRM Expressed Sequence) (Hermes).

GN RBPM5 OR HERMES.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=99195799; PubMed=1096065;

RA Gerber W.V., Yatskevych T.A., Antin P.B., Correia K.M., Conlon R.A.,

RT Kriegl P.A.;

RT "The RNA-binding protein gene, hermes, is expressed at high levels in

RT the developing heart."

RL Mech. Dev. 80:77-86(1999).

CC -1- FUNCTION: MAY BIND RNA.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING HEART.

CC -1- DEVELOPMENTAL STAGE: MRNA ALREADY DETECTED AT STAGE 7-8 IN THE

CC CARDIOGENIC MESODERM, AND BECOME ALMOST UNDETECTABLE IN THE OUTER

CC CURVATURE OF THE VENTRICULAR REGION WHEREAS REMAINING HIGH IN THE

CC DEVELOPING ATRIAL REGIONS.

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF129933; AAD30273.1; -

DR InterPro: IPR000504; RNA_rec_mot.

DR Pfam: PF00076; rrm; 1.

DR SMART: SM00360; RRM; 1.

DR PROSITE: PS50102; RRM; 1.

DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.

RN RNA-binding.

FT DOMAIN 22 99 RNA-BINDING (RRM).

FT DOMAIN 169 176 POLY-ALA.

SQ SEQUENCE 200 AA; 21856 MW; 4093B3C780BDC1DF CRC64;

Query Match 4.1%; Score 7; DB 1; Length 200;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 AAAAAL 14

Db 171 AAAAAL 177

RESULT 13

CTFL_MOUSE STANDARD; PRT; 203 AA.

AC Q60753;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Cardiotrophin-1 (CT-1).

GN CTFL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
MEDLINE=95166785; PubMed=7862649;
RA Pennica D., King K.L., Shaw K.J., Luis E., Rullamas J., Luch S.-M.,
RA Darbonne W.C., Knutson D.S., Yen R., Chien K.R., Baker J.B.,
RA Wood W.I.;
RT "Expression cloning of cardiotrophin 1, a cytokine that induces
RT cardiac myocyte hypertrophy." 92.1142-1146(1995).
RL Proc. Natl. Acad. Sci. U.S.A. 92.1142-1146(1995).
CC -1- FUNCTION: INDUCES CARDIAC MYOCYTE HYPERTROPHY IN VITRO. BINDS TO
CC AND ACTIVATES THE LEUKEMIA INHIBITORY FACTOR RECEPTOR (LIF
CC RECEPTOR)/GP 130 RECEPTOR COMPLEX.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE,
CC LIVER, LUNG AND KIDNEY. LOWER LEVELS IN TESTIS AND BRAIN. NO
CC EXPRESSION IN SPLEEN.
CC -1- SIMILARITY: BELONGS TO THE IL-6 SUPERFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: U18366; AAC52173.1; -
DR MGD: MGI:105115; Ctlf.
KW Cytokine.
SQ SEQUENCE 203 AA; 21509 MW; 8B3D414A0B3B232P CRC64;

Query Match 4.1%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALSVL 17
Db 91 AALSVL 97

RESULT 14
ID RL14-RAT STANDARD; PRT; 213 AA.
AC 063507;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L14.
GN RPL14.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40 AND 119-141.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96222520; PubMed=8670222;
RA Chan Y.-L., Olivera J., Wool I.G.;
RT "The primary structure of rat ribosomal protein L14."
RL Biochem. Biophys. Res. Commun. 222:427-431(1996).
CC -1- SIMILARITY: BELONGS TO THE L14E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: X94242; CAA63926.1; -
DR InterPro: IPR000302; K0W_motif.

InterPro: IPR002784; Ribosomal_L14e.
DR Pfam: PF00467; K0W; 1.
DR Pfam: PF01929; Ribosomal_L14e; 1.
KW Ribosomal protein; Repeat.
FT INTL_MET 0
FT DOMAIN 169 188 4 X 5 AA TANDEM REPEATS OF Q-K-A-[AS]-X.
FT REPEAT 169 173 1-1.
FT REPEAT 169 178 1-2.
FT REPEAT 174 183 1-3.
FT REPEAT 179 183 1-4.
FT REPEAT 184 188 1-4.
FT DOMAIN 192 197 2 X 3 AA TANDEM REPEATS OF K-G-Q.
FT REPEAT 192 194 2-1.
FT REPEAT 195 197 2-2.
SQ SEQUENCE 213 AA; 23207 MW; E14856D6070A1A44 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAANA 13
Db 149 IAAANA 155

RESULT 15
ID COBM_MYCTU STANDARD; PRT; 251 AA.
AC 010672;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Precorrin-4 C11-methyltransferase (EC 2.1.1.133) (Precorrin-3
DE methylase).
GN COBM OR RV2071C OR MT2131 OR MTCY49.10C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellton S., Squares S., Squares R.,
RA Sutcliffe J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence." Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Hart D., Hickey E., S.L.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METHYLTRANSFERASE THAT CATALYZES THE METHYLATION OF
CC C-11 IN PRECORRIN-4 TO FORM PRECORRIN-5.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-4 = S-
CC adenosyl-L-homocysteine + precorrin 5.
CC -1- PATHWAY: Cobalamin biosynthesis.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUWT, CYSG, CBIF/COBM
CC AND CBIL/COBI.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; 273966; CAA98218.1; -
DR EMBL; AE007063; AAK46411.1; ALT_INIT.
DR TIGR; MT2131; -
DR TubercuList; RV2071C; -
DR InterPro: IPR000878; Cor/por_Mettransf.
DR InterPro: IPR003043; Uropor_Mettransf.
DR Pfam: PF00590; TP_methylase; 1.
DR PROSITE; PS00839; SUMT_1; 1.
DR PROSITE; PS00840; SUMT_2; 1.
KW Cobalam biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 251 AA; 26420 MW; 8D8F21ABFC91CD3 CRC64;

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 251;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAAL 14
DB 114 AAAAAAAL 120

RESULT 16
ID GSHI_MOUSE STANDARD; PRT; 261 AA.
AC P31315;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein GSH-1.
GN GSHI OR GSH-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH Swiss;
RX MEDLINE=96172995; PubMed=8589431;
RA Valerius M.T., Li H., Stock J.L., Weinstein M., Kaur S., Singh G.,
RA Potter S.S.;
RT "Gsh-1: a novel murine homeobox gene expressed in the central nervous
RT system.";
RL Dev. Dyn. 203:337-351(1995).

RN [2]
RP SEQUENCE FROM N.A. AND FUNCTION.
RX MEDLINE=96181350; PubMed=8631293;
RA Li H., Zeitler P.S., Valerius M.T., Small K., Potter S.S.;
RT "Gsh-1, an orphan Hox gene, is required for normal pituitary
RT development.";
RL EMBO J. 15:714-724(1996).

RN [3]
RP SEQUENCE OF 146-205 FROM N.A.
RX MEDLINE=92073356; PubMed=1683707;
RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Potter S.S.;
RT "Identification of 10 murine homeobox genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).

CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT BINDS TO THE DNA
CC SEQUENCE 5'-GC[TA][AC]ATTA[GAI]-3', ACTIVATES THE TRANSCRIPTION OF
CC -1- SUBCELLULAR LOCATION: Nuclear
CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U21224; AAA96814.1; -
DR PIR; A37290; A37290.
DR PIR; A38809; A38809.
DR HSSP; P14653; 1B72.
DR MGD; MGI:95842; Gsh1.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00711; HOMEBOX_2; 1.
KW Transcription regulation; Activator; Homeobox; DNA-binding;
KW Nuclear protein; Developmental protein.
FT DOMAIN 110 117
FT DNA_BIND 146 205
SQ SEQUENCE 261 AA; 27727 MW; BA9D6B943594FF3 CRC64;

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAAL 14
DB 112 AAAAAAAL 118

RESULT 17
ID PYRK_METKA STANDARD; PRT; 262 AA.
AC P58887;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable dihydroorotate dehydrogenase electron transfer subunit.
GN PYRK OR MK0564.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Mal'kh A.G., Koonin E.V., Kozlovskii S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).

CC -1- FUNCTION: Is responsible for channeling the electrons from the
CC oxidation of dihydroorotate from the FMN redox center in the pyrd
CC submit to the ultimate electron acceptor NAD(+) (By similarity).
CC -1- COFACTOR: Binds a 2Fe-2S cluster and FAD (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC -1- SUBUNIT: Heterotetramer of 2 pyrk and 2 pyrd subunits (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE PYRK FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AE010350; AA001779.1; -
 DR PROSITE; PS00197; 2FE2S_FERRDOXIN; 1.
 KM Pyrimidine biosynthesis; Transport; Electron transport; Metal-binding;
 Itron; Iron-sulfur; Flapoprotein; FAD; Complete proteome.
 FT METAL 217 217 IRON-SULFUR 1 (2FE-2S) (BY SIMILARITY).
 FT METAL 222 222 IRON-SULFUR 1 (2FE-2S) (BY SIMILARITY).
 FT METAL 225 225 IRON-SULFUR 2 (2FE-2S) (BY SIMILARITY).
 FT METAL 234 234 IRON-SULFUR 2 (2FE-2S) (BY SIMILARITY).
 SQ SEQUENCE 262 AA; 28783 MW; 184A7DICE99B6761 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 262;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 TGVAPLV 53
 |||||
 Db 109 TGVAPLV 115

RESULT 18
 MTRC_METTH STANDARD; PRT; 266 AA.
 ID MTRC_METTH
 AC 027229;
 DT 15-JUL-1999 (rel. 38, Created)
 DT 15-JUL-1999 (rel. 38, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Tetrahydromethanopterin S-methyltransferase subunit C (EC 2.1.1.86)
 DE (N5-methyltetrahydromethanopterin--coenzyme M methyltransferase
 subunit C).
 GN MTRC OR MTH1161.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
 Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,
 RA Complete genome sequence of Methanobacterium thermoautotrophicum
 Delta: functional analysis and comparative genomics.*;
 RT J. Bacteriol. 179:7135-7155(1997).
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- FUNCTION: THIS ENZYME COMPLEX CATALYZES AN INTERMEDIATE STEP IN
 METHANOGENESIS, THE FORMATION OF METHYL-COENZYME M AND
 TETRAHYDROMETHANOPTERIN FROM COENZYME M AND N5-METHYL-
 TETRAHYDROMETHANOPTERIN. THIS IS A ENERGY-CONSERVING, SODIUM-ION
 TRANSLOCATING STEP (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 5-methyl-5,6,7,8-tetrahydromethanopterin + 2-
 mercaptoethanesulfonate = 5,6,7,8-tetrahydromethanopterin + 2-
 (methylthio)ethanesulfonate.
 CC -1- PATHWAY: Methanogenesis.
 CC -1- SUBUNIT: COMPOSED OF 8 DIFFERENT SUBUNITS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE000885; AAB85650.1; -
 DR TIGRPMAS; TIGR01148; mtrc; 1.
 KW Complete proteome; Transmembrane; Methanogenesis;

FT INTL MET 0 0 BY SIMILARITY.
 FT TRANSMEM 17 37 POTENTIAL.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 75 95 POTENTIAL.
 FT TRANSMEM 98 118 POTENTIAL.
 FT TRANSMEM 137 157 POTENTIAL.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 208 228 POTENTIAL.
 FT TRANSMEM 229 249 POTENTIAL.
 SQ SEQUENCE 266 AA; 26843 MW; 39B44F59344CDB9F CRC64;

Query Match 4.1%; Score 7; DB 1; Length 266;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSYL 17
 |||||
 Db 140 AALSYL 146

RESULT 19
 MTRC_METTH STANDARD; PRT; 266 AA.
 ID MTRC_METTH
 AC P80185; Q59583;
 DT 01-JUL-1993 (rel. 26, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Tetrahydromethanopterin S-methyltransferase subunit C (EC 2.1.1.86)
 DE (N5-methyltetrahydromethanopterin--coenzyme M methyltransferase
 subunit C).
 GN MTRC.
 OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=79929;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94039026; PubMed=8223548;
 RA Stupperich E., Juza A., Hoppert M., Mayer F.;
 RT Cloning, sequencing and immunological characterization of the
 RT cDNA-encoding subunit of the
 RT N5-methyltetrahydromethanopterin: coenzyme-M methyltransferase from
 RT Methanobacterium thermoautotrophicum.*;
 RL Eur. J. Biochem. 217:115-121(1993).
 RN [2]
 RP SEQUENCE OF 1-30.
 RX MEDLINE=9328732; PubMed=8477726;
 RA Geertner P., Ecker A., Fischer R., Linder D., Fuchs G., Thauer R.K.;
 RT Purification and properties of N5-
 RT methyltetrahydromethanopterin:coenzyme M methyltransferase from
 RT Methanobacterium thermoautotrophicum.*;
 RL Eur. J. Biochem. 213:537-545(1993).
 CC -1- FUNCTION: THIS ENZYME COMPLEX CATALYZES AN INTERMEDIATE STEP IN
 METHANOGENESIS, THE FORMATION OF METHYL-COENZYME M AND
 TETRAHYDROMETHANOPTERIN FROM COENZYME M AND N5-METHYL-
 TETRAHYDROMETHANOPTERIN. THIS IS A ENERGY-CONSERVING, SODIUM-ION
 TRANSLOCATING STEP.
 CC -1- CATALYTIC ACTIVITY: 5-methyl-5,6,7,8-tetrahydromethanopterin + 2-
 mercaptoethanesulfonate = 5,6,7,8-tetrahydromethanopterin + 2-
 (methylthio)ethanesulfonate.
 CC -1- PATHWAY: Methanogenesis.
 CC -1- SUBUNIT: COMPOSED OF 8 DIFFERENT SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; X73123; CAA51554.1; -
 DR

DR PIR, S30343, S30343.
 DR TIGR: TIGR0148; mtic; 1.
 KW Transferase; Methyltransferase; Transmembrane; Methanogenesis.
 FT INIT MET 0 0
 FT TRANS MEM 17 37 POTENTIAL.
 FT TRANS MEM 38 58 POTENTIAL.
 FT TRANS MEM 75 95 POTENTIAL.
 FT TRANS MEM 98 118 POTENTIAL.
 FT TRANS MEM 137 157 POTENTIAL.
 FT TRANS MEM 162 182 POTENTIAL.
 FT TRANS MEM 208 228 POTENTIAL.
 FT TRANS MEM 229 249 POTENTIAL.
 FT CONFLICT 30 30 Y -> S (IN REF. 2).
 SQ SEQUENCE 266 AA; 26986 MW; 2984360328207807 CRC64;

Query Match
 Best Local Similarity 4.1%; Score 7; DB 1; Length 266;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALSVL 17
 |||||
 Db 140 AAALSVL 146

RESULT 20
 HXDB_HUMAN STANDARD; PRT; 290 AA.
 AC P13378;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-D8 (Hox-4E) (Hox-5.4).
 GN HOXD8 OR HOXA4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Birren B., Linton L., Nusbaum C., Lander E.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Kosaki K., Kosaki R., Suzuki T., Yoshitashi H., Sasaki K., Matsuo N.;
 RL "A complete mutation analysis panel of human HOX genes";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE OF 193-287 FROM N.A.
 RX MEDLINE=89306602; PubMed=2568311;
 RA Oliver G., Sidell N., Fiske N., Heinzmann C., Mohandas T.,
 RA Sparkes R.S., de Robertis E.M.;
 RL "Complementary homeo protein gradients in developing limb buds.";
 RL Genes Dev. 3:641-650(1989).
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC -1- SUBCELLULAR LOCATION: Nucleus.
 CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AC009336; -; NOT ANNOTATED_CDS.
 DR EMBL: AY014304; AA642152.1;
 DR EMBL: AY014303; AA642152.1; JOINED.
 DR EMBL: X15507; CAA33529.1;
 DR PIR: B32830; B32830.
 DR PIR: S05957; S05957.

DR HSP: P02833; 9ANT.
 DR TRANSFAC; T03332;
 DR Gene; HGNC:5139; HOXD8.
 DR MIM; 142985;
 DR InterPro; IPR001827; Antennapedia.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00025; ANTENNAPEIDIA.
 DR PRODOM; PD000010; Homeobox.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00032; HOMEBOX_1; 1.
 DR PROSITE; PS00032; ANTENNAPEIDIA; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 15 23 POLY-ALA.
 FT DOMAIN 45 50 POLY-ALA.
 FT DOMAIN 109 123 POLY-PRO.
 FT DNA_BIND 197 256 HOMEBOX.
 FT CONFLICT 287 287 G -> A (IN REF. 3).
 SQ SEQUENCE 290 AA; 31910 MW; 75FF95A73E2A85F CRC64;

Query Match
 Best Local Similarity 4.1%; Score 7; DB 1; Length 290;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
 |||||
 Db 45 AAAAAL 51

RESULT 21
 PSBO_CHURE STANDARD; PRT; 291 AA.
 AC P12853;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Oxygen-evolving enhancer protein 1, chloroplast precursor (OEE1).
 GN PSBO.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=137C / CC-125;
 RA Mayfield S.P., Schlimer-Rahire G., Frank H., Zuber H., Rochaix J.-D.;
 RL "Analysis of the genes of the OEE1 and OEE3 proteins of the
 RL Photosystem II complex of Chlamydomonas reinhardtii.";
 RL Plant Mol. Biol. 12:683-693(1989).
 CC -1- FUNCTION: STABILIZES THE MANGANESE CLUSTER WHICH IS THE PRIMARY
 CC SITE OF WATER SPLITTING (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane; associated
 CC with the photosystem II complex.
 CC -1- SIMILARITY: BELONGS TO THE PSBO FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X13826; CAA32053.1;
 DR PIR: S05508; S05508.
 DR InterPro; IPR002628; PSII_MSP.
 DR Pfam; PF01716; MSP; 1.
 KW Photosynthesis; Photosystem II; Chloroplast; Transit peptide;
 KW Thylakoid; Membrane; Manganese.
 FT TRANSIT 1 52 CHLOROPLAST.

```

FT CHAIN 53 291 OXYGEN-EVOLVING ENHANCER PROTEIN 1.
SQ SEQUENCE 291 AA; 30522 MW; 5DF4DFAB1324267 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
   |||||
Db 34 AAAAAL 40

RESULT 22
RBSK_BACHD STANDARD; PRT; 294 AA.
AC O9K6K1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribokinase (EC 2.7.1.15).
GN RBSK OR BH3728.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 /JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: ATP + D-ribose -> ADP + D-ribose 5-phosphate.
CC -1- PATHWAY: Ribose metabolism; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: AP001519; BAB07447.1; -.
CC HSSP: P05054; 1RK2.
CC InterPro: IPR002173; PfKB.
CC InterPro: IPR002139; Ribokinase.
CC Pfam: PF00294; PfKB; 1.
CC PRINTS: PR00990; RIBOKINASE.
CC PROSITE: PS00583; PFKB_KINASES_1; 1.
CC PROSITE: PS00584; PFKB_KINASES_2; 1.
CC Transfaser: Kinase; Complete proteome.
SQ SEQUENCE 294 AA; 31089 MW; 8C13E0FCF5E89FDE CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAALSV 16
   |||||
Db 264 AAAALSV 270

RESULT 23
GSH2_HUMAN STANDARD; PRT; 304 AA.
AC G9BZM3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein GSH-2.
GN GSH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Sakai T., Sakamoto S., Nakamura K., Muraki T.;
RT "Human homeobox protein GSH-2.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cools J., Marynen P.;
RT "The sequence of the human GSH2 gene.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT BINDS TO THE DNA
CC SEQUENCE 5'-CNAATTAG-3' (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: AB028638; BAB84822.1; -.
CC EMBL: AF306344; AAK00880.1; -.
CC EMBL: AF306343; AAK00880.1; JOINED.
CC HSSP: P14653; 1B72.
CC InterPro: IPR000047; HTH_repressr.
CC InterPro: IPR001356; Homeobox.
CC Pfam: PF00046; homeobox; 1.
CC PRINTS: PR00024; HOMEBOX.
CC PRINTS: PR00031; HTHREPRESSR.
CC PRODOM: PD000010; Homeobox; 1.
CC SMART: SM00389; HOX; 1.
CC PROSITE: PS00027; HOMEBOX_1; 1.
CC PROSITE: PS00027; HOMEBOX_2; 1.
CC Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
CC Developmental protein.
CC DNA_BIND 202 261 HOMEBOX.
CC DOMAIN 124 130 POLY-HIS.
CC DOMAIN 134 139 POLY-HIS.
CC DOMAIN 147 162 POLY-ALA.
SQ SEQUENCE 304 AA; 32061 MW; E896D542248E6C1 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
   |||||
Db 157 AAAAAL 163

RESULT 24
GSH2_MOUSE STANDARD; PRT; 305 AA.
AC P31316;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein GSH-2.
GN GSH2 OR GSH-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIH SWISS;
RX MEDLINE=95344993; PubMed=7619729;
RA Hsieh-Li H.M., Witte D.P., Stancsik J.C., Weinstein M., Li H.,
RT Potter S.S.;
RL "Gsh-2, a murine homeobox gene expressed in the developing brain.";
RM Mech. Dev. 50:177-186(1995).
[2]
RP SEQUENCE OF 203-262 FROM N.A.
RX MEDLINE=9207356; PubMed=1683707;
RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Potter S.S.;
RT "Identification of 10 murine homeobox genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT BINDS TO THE DNA
CC SEQUENCE 5'-CNAATTAG-3'.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING BRAIN.
CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S79041; AAB34947.1; -
DR PIR: B37290; B37290.
DR PIR: B38809; B38809.
DR HSSP: P14653; 1B72.
DR MGD: MGT:95843; Gsh2.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox.1.
DR PRINTS: PR00024; HOMEBOX.
DR ProDom: PD00010; HOMEBOX.1.
DR SMART: SM00389; HOX.1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KW Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW Developmental protein.
FT DNA_BIND 203 262 HOMEBOX.
FT DOMAIN 124 130 POLY-HIS.
FT DOMAIN 134 139 POLY-HIS.
FT DOMAIN 147 163 POLY-ALA.
SQ SEQUENCE 305 AA; 32167 MW; 51E7F2DB76E32608 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 305;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 158 AAAAAA 164

```

```

RC STRAIN-DSM 1411;
RX MEDLINE=93021102; PubMed=1404376;
RA Engert C., Krueger K., Offner S., Pfeiffer F.;
RT "Three different but related gene clusters encoding gas vesicles in
RT halophilic archaea.";
RL J. Mol. Biol. 227:586-592(1992).
CC -1- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS
CC VESICLE SYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64701; CAA45954.1; -
DR PIR: S28125; S28125.
DR Gas vesicle.
SQ SEQUENCE 322 AA; 36382 MW; EEDB60511E10EE5 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 322;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 164 ONRRAL 170
DB 234 ONRRAL 240

```

```

RESULT 26
ID SLP1_DROME STANDARD; PRT; 322 AA.
AC P32030; O9YOV4;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fork head domain transcription factor slp1 (Slippy paired locus
DE Protein 1).
GN SLP1 OR FDB OR CG16738.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Canton-S, and Oregon-R;
RX MEDLINE=92275347; PubMed=1317319;
RA Grossniklaus U., Pearson R.K., Gehring W.J.;
RT "The Drosophila sloppy paired locus encodes two proteins involved in
RT segmentation that show homology to mammalian transcription factors.";
RL Genes Dev. 6:1030-1051(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Chang M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Boltskov S.,
RA Borokova D., Botchan M.R., Bouck B., Brokstein P., Brotter P.,
RA Burrus K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,

```

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gloeck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.H., Ilegan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenilson J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., Mepherston D.,
 RA Mervinov G., Mishina N.Y., Mobarry C., Morris J., Moshneff A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard D.R., Paele J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.M., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN SEGMENTATION. REQUIRED
 CC FOR THE FORMATION OF THE MANDIBULAR LOBE. DIFFERENT LEVELS OF SLIP
 CC ACTIVITY SEEM TO BE REQUIRED IN DIFFERENT SEGMENTS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE POSTERIOR HALF OF EACH
 CC PARASEGMENT JUST ANTERIOR TO THE PARASEGMENTAL BOUNDARY.
 CC -1- DEVELOPMENTAL STAGE: PRESENT AT 0-3 HRS OF EMBRYOGENESIS. MAXIMAL
 CC EXPRESSION AT 3-6 HRS. STRONG RE-EXPRESSION IN FIRST-INSTAR
 CC LARVAE.
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X66095; CAA46889.1; -;
 DR EMBL: X66096; CAA46890.1; -;
 DR EMBL: AE003578; AAF51058.1; -;
 DR PIR: S23053; S23053.
 DR PIR: S23054; S23054.
 DR HSSP: Q63245; 2HRH.
 DR TRANSFAC: T01055; -;
 DR FLYBASE: FBgn0003430; slp1.
 DR InterPro: IPR001766; TF_Fork_head.
 DR Pfam: PF00250; Fork_head; 1.
 DR PRINTS: PR00053; FORKHEAD.
 DR PRODOM: PD000425; TF_Fork_head; 1.
 DR SMART: SM00339; FH; 1.
 DR PROSITE: PS00657; FORK_HEAD_1; 1.
 DR PROSITE: PS00658; FORK_HEAD_2; 1.
 DR PROSITE: PS00659; FORK_HEAD_3; 1.
 DR DNA-binding, Developmental protein: Nuclear protein;
 KW Transcription regulation; Pair-rule protein; Polymorphism.
 FT DNA_BIND 119 210 FORK-HEAD.
 FT DOMAIN 256 262 POLY-ALA.
 FT VARIANT 246 246 P -> S (IN STRAIN OREGON-R AND BERKELEY).
 FT VARIANT 303 303 P -> Q (IN STRAIN OREGON-R AND BERKELEY).
 FT SEQUENCE 322 AA; 36202 MW; 24CDE9F0102024C4 CRC64;
 SQ
 Query Match 4.1%; Score 7; DB 1; Length 322;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 27
 A85B_MYCKA STANDARD; PRT; 325 AA.
 ID A85B_MYCKA
 AC P21160;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antigen 85-B precursor (85B) (Extracellular alpha-antigen) (Antigen 85
 DE complex B) (Ag85B) (Mycolyl transferase 85B) (Ec 2.3.1.-)
 DE (F18pneocin-binding protein B).
 GN F18pneocin-binding protein B.
 OS Mycobacterium kansasii.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxID=178;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90129315; PubMed=2404875;
 RA Matsuo K., Yamaguchi R., Yamazaki A., Tasaka H., Terasaka K.,
 RA Yamada T.;
 RT "Cloning and expression of the gene for the cross-reactive alpha
 RL antigen of *Mycobacterium kansasii*.";
 RL Infect. Immun. 58:550-556(1990).
 CC -1- FUNCTION: PROTEINS OF THE ANTIGEN 85 COMPLEX ARE RESPONSIBLE FOR
 CC THE HIGH AFFINITY OF MYCOBACTERIA TO FIBRONECTIN. POSSESSES A
 CC MYCOLYTIC TRANSFERASE ACTIVITY REQUIRED FOR THE BIOSYNTHESIS OF
 CC TREHALOSE DIMYCOLATE (CORD FACTOR), A DOMINANT STRUCTURE NECESSARY
 CC FOR MAINTAINING CELL WALL INTEGRITY (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: TO OTHER MYCOBACTERIUM A85 ANTIGENS AND TO THE
 CC N-TERMINAL OF C. GLUTAMICUM PSI PROTEIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X53897; CAA37868.1; -;
 DR PIR: A37185; A37185.
 DR HSSP: P31953; IDOY.
 DR InterPro: IPR000801; Esterase-put.
 DR Pfam: PF00756; Esterase; 1.
 KW Transferase; Acyltransferase; Antigen; Signal.
 FT SIGNAL 1 40
 FT CHAIN 41 325
 FT ACT_SITE 166 166 ANTIGEN 85-B.
 FT ACT_SITE 270 270 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 FT SEQUENCE 325 AA; 34323 MW; 5F2281BC48AE30D CRC64;
 SQ
 Query Match 4.1%; Score 7; DB 1; Length 325;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 DB 257 AAAAAA 263

RESULT 28
 GDC_BOVIN STANDARD; PRT; 330 AA.
 ID GDC_BOVIN
 AC Q01888;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Grave's disease carrier protein (GDC) (Mitochondrial solute carrier
 DE protein homology).

GN SIC25A16 OR GDA OR GDC.
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=93091248; PubMed=1457817;
 RA Flernonte G., Runswick M.J., Walker J.E., Palmieri F.;
 RT "Sequence and pattern of expression of a bovine homologue of a human
 mitochondrial transport protein associated with Grave's disease.";
 RL DNA Seq. 3:71-78(1992).
 CC
 CC -1- FUNCTION: Required for the accumulation of coenzyme A in the
 CC mitochondrial matrix (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (By similarity).
 CC -1- TISSUE SPECIFICITY: MOSTLY IN THYROID, LIVER, LUNG, KIDNEY AND
 CC TO A LESSER EXTENT IN HEART AND SKELETAL MUSCLE.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X66035; CAA46834.1; -
 DR PIR: S26596; S26596;
 DR InterPro: IPR002067; Mit-carrier.
 DR InterPro: IPR001993; Mitoch-carrier.
 DR Pfam: PF00153; mito_carr; 3.
 DR PRINTS: PR00926; MITOCHCARRIER.
 DR PROSITE: PS00215; MITOCH_CARRIER; 3.
 DR Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.
 KW TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 189 209 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT REPEAT 1 121 1.
 FT REPEAT 122 217 2.
 FT REPEAT 218 330 3.
 SQ SEQUENCE 330 AA; 36085 MW; 4C614701D2B8DEA9 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 330;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 AAAAAA 14
 Db 2 AAAAAA 8
 RESULT 29
 ID RLAO_METUA STANDARD; PRT; 338 AA.
 AC P34049;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Acidic ribosomal protein P0 homolog (L10E).
 GN RPLP0 OR M05059.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 CC NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RL Science 273:1058-1073(1996).
 CC
 CC -1- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
 CC OF E-COLI PROTEIN L10.
 CC
 CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U67500; AAB8449.1; -
 DR TIGR: M05059;
 DR InterPro: IPR001790; Ribosomal_L10.
 DR Pfam: PF00466; Ribosomal_L10; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 338 AA; 36751 MW; 63A6ARD357E3052D CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 338;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 61 ALASKLP 67
 Db 282 ALASKLP 288
 RESULT 30
 ID GUB_FIBSU STANDARD; PRT; 349 AA.
 AC P17989;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Beta-glucanase precursor (EC 3.2.1.73) (Endo-beta-1,3-1,4 glucanase)
 DE (1,3-1,4-beta-D-glucan 4-glucanohydrolase) (Mixed linkage beta-
 DE glucanase) (Lichenase).
 OS Fibrobacter succinogenes (Bacteroides succinogenes).
 CC Bacteria; Fibrobacter/acidobacteria group; Fibrobacter group;
 CC Fibrobacter.
 CC NCBI_TaxID=833;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.
 RC STRAIN=Isolate S85;
 RX MEDLINE=90299807; PubMed=2193918;
 RA Teather R.M., Ertle J.D.;
 RT "DNA sequence of a Fibrobacter succinogenes mixed-linkage
 beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";
 RL J. Bacteriol. 172:3837-3841(1990).
 CC
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in beta-D-glucans containing 1,3- and 1,4-bonds.
 CC
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M33676; AAA24896.1; -
 DR PIR; A44507; A44507.
 DR HSSP; P23904; 1A4K.
 DR InterPro; IPR000757; glyco_hydro_16.
 DR Pfam; PF00722; glyco_hydro_16; 1.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolase; Glycosidase; Signal; Repeat.
 FT SIGNAL 1 27
 FT CHAIN 28 349
 FT ACT_SITE 79 79 BETA-GLUCANASE.
 FT ACT_SITE 83 83 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 271 307 PROTON DONOR (BY SIMILARITY).
 FT REPEAT 271 277 5 X 7 AA TANDEM REPEATS OF P-X-S-S-S-S-X.
 FT REPEAT 278 284 1.
 FT REPEAT 285 291 2.
 FT REPEAT 292 298 3.
 FT REPEAT 301 307 4.
 FT REPEAT 307 307 5.
 SQ SEQUENCE 349 AA; 37737 MW; 16DC4F5BDEFC578A CRC64;

Query Match 4.1%; Score 7; DB 1; Length 349;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
 111111
 Db 15 AAAAAL 21

RESULT 31
 DKK3_HUMAN STANDARD; PRT; 350 AA.
 AC Q9UBP4; Q9UBP4; 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dickkopf related protein-3 precursor (DKK-3) (Dickkopf-3) (hdkk-3).
 GN DKK3 OR REIC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=fetal brain;
 RX MEDLINE=20035735; PubMed=10570958; Robinson K., Chickering T.W.,
 RA Kirupak V.E., Sharp J.D., Jiang C., Robison K., Chickering T.W.,
 RA Amaravadi L., Brown D.E., Guyot D., Mays G., Leiby K., Chang B.,
 RA Duong T., Goodearl A.D.J., Gearing D.P., Sokol S.Y., McCarthy S.A.;
 RT "Functional and structural diversity of the human Dickkopf gene
 family";
 RT Gene 238:301-313(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RA Tanaka S., Sugimachi K., Sugimachi K.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20119095; PubMed=10652205;
 RA Tsuji T., Miyazaki M., Sakaguchi M., Inoue Y., Namba M.;
 RT "A REIC gene shows down-regulation in human immortalized cells and
 human tumor-derived cell lines";
 RL Biochem. Biophys. Res. Commun. 268:20-24(2000).
 [4]
 RP SEQUENCE FROM N.A.
 RA Tate G., Mitsuya T.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE FROM N.A.
 RC PubMed=11814687;
 RA Kobayashi K., Ouchida M., Tsuji T., Hanafusa H., Miyazaki M.,
 RA Namba M., Shimizu N., Shimizu K.;
 RT "Reduced expression of the REIC/Dkk-3 gene by promoter-
 hypermethylation in human tumor cells";

RL Gene 282:151-158(2002).
 [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Straube R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INHIBITOR OF WNT SIGNALING PATHWAY (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN HEART, BRAIN, AND SPINAL
 CC CORD.
 CC -1- PTM: N-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE DICKKOPF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF177396; AAF02676.1; -
 DR EMBL; AB033421; BAA85488.1; -
 DR EMBL; AB034203; BAA80548.1; -
 DR EMBL; AB035182; BAA87044.2; -
 DR EMBL; AB045205; BAA87044.2; JOINED.
 DR EMBL; AB045206; BAA87044.2; JOINED.
 DR EMBL; AB045207; BAA87044.2; JOINED.
 DR EMBL; AB045208; BAA87044.2; JOINED.
 DR EMBL; AB045209; BAA87044.2; JOINED.
 DR EMBL; AB045210; BAA87044.2; JOINED.
 DR EMBL; AB057591; BAA84360.1; -
 DR EMBL; AB057804; BAA84361.1; -
 DR EMBL; BC007660; AAO07660.1; -
 DR Gene; HGNC:2893; DKK3.
 DR MIM; 605416; -
 KW Developmental protein; Signal; Glycoprotein.
 FT SIGNAL 1 16
 FT CHAIN 17 350
 FT DOMAIN 147 195 DICKKOPF RELATED PROTEIN-3.
 FT DOMAIN 208 284 DKK-TYPE CYS-2.
 FT DOMAIN 338 343 POLY-ALA.
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 335 335 G -> R (IN REF. 4).
 SQ SEQUENCE 350 AA; 38291 MW; 72F504122B40AFEE CRC64;

Query Match 4.1%; Score 7; DB 1; Length 350;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
 111111
 Db 338 AAAAAL 344

RESULT 32
 KLF2_MOUSE STANDARD; PRT; 354 AA.
 ID KLF2_MOUSE
 AC O60843;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Kruppel-like factor 2 (Lung Kruppel-like factor).
 GN KLF2 OR IKLF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.

CC STRAIN-C57BL/6 X CBA; TISSUE=Lung;
 RA MEDLINE=96025976; PubMed=7565748;
 RT Anderson K.P., Kern C.B., Crabbe S.C., Lingrel J.B.;
 RT "Isolation of a gene encoding a functional zinc finger protein
 RT homologous to erythroid Kruppel-like factor: identification of a new
 RT multigene family.";
 RL Mol. Cell. Biol. 15:5957-5965(1995).
 CC -1- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER
 CC AND ACTIVATES TRANSCRIPTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN THE LUNGS AND
 CC SPLEEN.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U25096; AAA86728.1; -
 DR HSSP: P08047; ISP2.
 DR TRANSFAC: T01677; -
 DR MGD: MGI:134272; Klf2.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf-C2H2; 3.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR PRODOM: PD000003; Znf_C2H2; 2.
 DR SMART: SM00355; Znf_C2H2; 3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 3.
 DR Transcription regulation: Activator; Zinc-finger; Metal-binding;
 DR DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 60 70 POLY-PRO.
 FT 165 169 POLY-PRO.
 FT 224 230 POLY-ALA.
 FT 271 353 ZINC_FINGERS.
 FT 271 295 C2H2-TYPE.
 FT 301 325 C2H2-TYPE.
 FT 331 353 C2H2-TYPE.
 SQ SEQUENCE 354 AA; 37700 MW; CAA99D018AC5BAF7 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 354;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 AAAAAL 14
 DB 225 AAAAAL 231
 RESULT 33
 KLF2_HUMAN STANDARD; PRT; 355 AA.
 AC Q9Y5W3; Q9UKR6; Q9U5S5;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Kruppel-like factor 2 (Lung Kruppel-like factor).
 GN KLF2 OR KLF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA MEDLINE=99231781; PubMed=10217429;
 RA Kozyrev S.V., Hansen L.L., Poltarau A.B., Dominsky D.A.,
 RA Kisselev L.L.;

RT "Structure of the human CpG-island-containing lung Kruppel-like factor
 RT (KLF2) gene and its location in chromosome 19p13.11-13 locus.";
 RL FEBS Lett. 448:149-152(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99389728; PubMed=10458913;
 RA Wanl M.A., Conkright M.D., Jeffries S., Hughes M.J., Lingrel J.B.;
 RT "cDNA isolation, genomic structure, regulation, and chromosomal
 RT localization of human lung Kruppel-like factor.";
 RL Genomics 60:78-86(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lee H.J., Kim M.K., Kim Y.H., Seo J.M., Lee H.M., Chung H.J.,
 RA Sohn M.Y., Hwang S.Y., Im S.O., Jung E.J., Kim J.C.;
 RT "A catalogue of genes in the human dermal papilla cells as identified
 RT by expressed sequence tags.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER
 CC AND ACTIVATES TRANSCRIPTION (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF123344; AAD25076.1; -
 DR EMBL: AF134053; AAD55891.1; -
 DR EMBL: AF205849; AAF13295.1; -
 DR HSSP: P08047; ISP2.
 DR TRANSFAC: T04958; -
 DR GeneW: HGNC:6347; KLF2.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf-C2H2; 3.
 DR PRODOM: PD000003; Znf_C2H2; 2.
 DR SMART: SM00355; Znf_C2H2; 3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 3.
 DR Transcription regulation: Activator; Zinc-finger; Metal-binding;
 DR DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 62 71 POLY-PRO.
 FT 130 135 POLY-GLY.
 FT 167 171 POLY-PRO.
 FT 225 231 POLY-ALA.
 FT 272 354 ZINC_FINGERS.
 FT 272 296 C2H2-TYPE.
 FT 302 326 C2H2-TYPE.
 FT 332 354 C2H2-TYPE.
 FT 43 43 S -> N (IN REF. 2).
 FT 104 104 L -> P (IN REF. 1).
 FT 175 175 P -> S (IN REF. 2).
 FT 184 184 L -> M (IN REF. 2).
 SQ SEQUENCE 355 AA; 37419 MW; D5849C831D676AE1 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 355;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 AAAAAL 14
 DB 226 AAAAAL 232
 RESULT 34
 BET3_MESAU STANDARD; PRT; 367 AA.
 ID BET3_MESAU

AC 009029; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE BERR3 protein.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96140430; PubMed=8552091;
 RA Peyton M., Stellrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,
 RA Tsai M.-J.;
 RA "BERR3, a novel helix-loop-helix protein, can act as a negative
 RT regulator of BERR2 and MyoD-responsive genes.";
 RT Mol. Cell. Biol. 16:626-633(1996).
 CC -1- FUNCTION: INHIBITS DNA BINDING OF TCF3 (E47) HOMODIMERS AND TCF3
 CC (E47) / NEUROD1 HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF
 CC NEUROD1 AND MYOD-RESPONSIVE GENES, PROBABLY BY HETERODIMERIZATION
 CC WITH CLASS A BASIC HELIX-LOOP-HELIX FACTORS. DESPITE THE PRESENCE
 CC OF AN INTACT BASIC DOMAIN, DOES NOT BIND TO DNA.
 CC -1- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).
 CC -1- TISSUE SPECIFICITY: KIDNEY, LUNG, BRAIN AND PANCREAS (INSULINOMA).
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "ATOMAL" SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; S80870; AAB50691.1; -;
 DR INTERPRO; IPR001092; HLH_basic.
 DR Pfam; PF00010; HLH; 1.
 DR SMART; SM00353; HLH; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 DR PROSITE; PSS0888; HLH_2; 1.
 KW Nuclear protein; transcription regulation; Repressor.
 FT DOMAIN 11 14 POLY-ALA.
 FT DOMAIN 58 62 POLY-SER.
 FT DOMAIN 83 99 POLY-GLY.
 FT DOMAIN 174 179 POLY-GLY.
 FT DOMAIN 204 217 POLY-GLY.
 FT DNA_BIND 229 240 BASIC DOMAIN.
 FT DOMAIN 241 282 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 311 319 POLY-ALA.
 SQ SEQUENCE 367 AA; 35905 MW; 6CAB9AF96E85F77 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 367;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kumano M., Tamakoshi A., Yamane K.;
 RA "A 32 kb nucleotide sequence from the region of the *lincomycin*-
 RT resistance gene (22-25 degree) of the *Bacillus subtilis* chromosome and
 RT identification of the site of the *lin-2* mutation."
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBD databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Berteiro M.G., Bessieres P., Bolotin A., Borczyk S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Enlian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Goldthly E.J., Grandi G.,
 RA Gutseppl G., Guy B.J., Haga K., Haeckel H., Hall M.F., Itaya M., Jones L.,
 RA Hilbert H., Holsappel S., Hosono S., Klier-Blanchard M., Klein C.,
 RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidis A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Meade N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadleir Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Torkel M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
 RA Tosato V., Uchlyama S., Vandenbol M., Vannier F., Vassartol A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weissmeyer T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*."
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE OF 1-319 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95219079; PubMed=7704254;
 RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;
 RA "Determination of a 21548 bp nucleotide sequence around the 24
 RT degrees region of the *Bacillus subtilis* chromosome."
 RL Microbiology 141:269-275(1995).
 RN [4]
 RP SEQUENCE OF 1-147 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=92339527; PubMed=1353026;
 RA Awade A., Cleuzat P., Gonzales T., Robert-Baudouy J.;
 RA "Characterization of the *pcp* gene encoding the pyrrolidone carboxyl
 RT peptidase of *Bacillus subtilis*."
 RL FEBS Lett. 305:67-73(1992).
 CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to two
 CC frameshifts.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch.

CC EMBL: AB000617; BAA22227.1; -
CC EMBL: 299105; CAB12060.1; -
DR EMBL: D30808; -; NOT_ANNOTATED_CDS.
DR EMBL: X66034; -; NOT_ANNOTATED_CDS.
DR Subtilist: Bg11176; ycbu.
DR InterPro: IPR00192; AminoTransfV.
DR Pfam: PF00266; aminotran_5.1.
DR PROSITE: PS00595; AA_TRANSF_5.1.
KW Hypothetical protein; Signal; Transferase; AminoTransferase;
KW Pyridoxal phosphate; Complete proteome.
FT SIGNAL 1 30
FT CHAIN 1 370
FT BINDING 207 207
FT CONFLICT 92 92
FT CONFLICT 133 133
FT CONFLICT 319 319
SQ SEQUENCE 370 AA; 40809 MM; 38C91AF0FC27DF53 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 370;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 AAAAALS 15
Db 265 AAAAALS 271

RESULT 36
FL_ORYSA STANDARD; PRT; 389 AA.
AC 024175;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative transcription factor FL (REFL).
GN FL.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Toride; TISSUE=Panicle;
RA MEDLINE=98151445; PubMed=9482818;
KY Kyojuka J., Konishi S., Nemoto K., Izawa T., Shimamoto K.;
FT "Down-regulation of RFL, the FLO/LFY homolog of rice, accompanied
FT with panicle branch initiation.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1979-1982(1998).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: In very young panicle but not in mature
CC florets, mature leaves, roots or apical meristems.
CC -1- SIMILARITY: BELONGS TO THE FLO / LFY FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AB005620; BAA21547.1; -
DR InterPro: IPR002910; FLO_LFY; 1.
DR Pfam: PF01698; FLO_LFY; 1.
KW Transcription regulation; Activator; DNA-binding;
KW Nuclear protein; Developmental protein.
FT DOMAIN 22 37
FT DOMAIN 201 204
FT DOMAIN 213 217
POLY-PRO.
POLY-GLY.
POLY-GLY.

FT DOMAIN 378 383
SQ SEQUENCE 389 AA; 42530 MM; 5132AA6A34B0868A CRC64;
Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 389;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
Db 378 AAAAAL 384

RESULT 37
HB9_HUMAN STANDARD; PRT; 401 AA.
AC P50219;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein HB9.
GN HLXB9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA MEDLINE=94327547; PubMed=7914194;
KY Harrison K.A., Druey K.M., Deguchi Y., Tuscano J.M., Kehrl J.H.;
FT "A novel human homeobox gene distantly related to proboscipedia is
FT expressed in lymphoid and pancreatic tissues.";
RL J. Biol. Chem. 269:19968-19975(1994).
CC -1- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND PANCREATIC TISSUES.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U07664; AAB60647.1; -
DR EMBL: U07663; AAB60647.1; JOINED.
DR HSSP: P14653; 1B72.
DR TRANSFAC: T03420; -
DR GeneW: HGNC:4979; HLXB9.
DR MIM: 142994; -
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEBOX.
DR PRODOM: PD000010; Homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein; Transcription regulation.

FT DOMAIN 39 48
FT DOMAIN 97 111
FT DOMAIN 120 135
FT DOMAIN 169 177
FT DNA BIND 242 301
FT DOMAIN 316 325
SQ SEQUENCE 401 AA; 40932 MM; 0006AED71D594FE CRC64;
Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 401;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14

DB 170 AAAAAL 176

RESULT 38
NU4M_ASCSU STANDARD: PRT: 409 AA.
ID NU4M_ASCSU
AC P24880:
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
GN ND4.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC Mitochondrion.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN NCBI_TaxID=6253;
RP [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Body wall muscle, and Egg;
RA MEDLINE=92201635; PubMed=1551572;
RA Okimoro R., Macfarlane J.L., Clary D.O., Wolstenholme D.R.;
RT "The mitochondrial genomes of two nematodes, Caenorhabditis elegans
and Ascaris suum";
RT Genetics 130:471-498(1992).
RL -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X54253; CAA38170.1; -
DR PIR: S26021; S26021.
DR InterPro: IPR001750; Oxidored_g1.
DR Pfam: PF00361; Oxidored_g1.1
DR Oxidoreductase; NAD; ubiquinone; Mitochondrion.
KW Oxidoreductase; NAD; ubiquinone; NAD135C1CA558786 CRC64;
SO SEQUENCE 409 AA; 47064 MW; A4B135C1CA558786 CRC64;
Query Match 4.1%; Score 7; DB 1; Length 409;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 68 SLVYDF 74
DB 147 SLVYDF 153
RESULT 39
RL4_HUMAN STANDARD: PRT: 427 AA.
ID RL4_HUMAN
AC P36578; P39029; Q96929;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L4 (L1).
GN RPL4 OR RPL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Euteleostomi; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=94092742; PubMed=8268230;
RA Bagni C., Mariotti P., Annesi F., Amaldi F.;
RT "Human ribosomal protein L4: cloning and sequencing of the cDNA and
RT primary structure of the protein";
RT Biochim. Biophys. Acta 1216:475-478(1993).

RN [2]
RA REVISIONS.
RA Bagni C.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RC SEQUENCE FROM N.A.
RC TISSUE=Lymphoma;
RA Kato S.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RC SEQUENCE FROM N.A.
RC MEDLINE=21864036; PubMed=11875025;
RA Yoshinaka H., Uechi T., Asakawa S., Kawasaki K., Kato S., Higa S.,
RA Maeda N., Minoshima S., Tanaka T., Shimizu N., Kenmochi N.;
RT "The human ribosomal protein genes: sequencing and comparative
RT analysis of 73 genes";
RT Genome Res. 12:379-390(2002).
RN [5]
RC SEQUENCE FROM N.A.
RC TISSUE=Colon, Eye, Muscle, Pancreas, and Skin;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RC -1- SIMILARITY: BELONGS TO THE LAE FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X73974; CAA52154.1; -
DR EMBL: L20868; AAA60281.1; ALT_SEQ.
DR EMBL: D23660; BAA04887.1; -
DR EMBL: BC001365; AAH01365.1; -
DR EMBL: BC005817; AAH05817.1; -
DR EMBL: BC007259; AAH07259.1; -
DR EMBL: BC007748; AAH07748.1; -
DR EMBL: BC007996; AAH07996.1; -
DR EMBL: BC009888; AAH09888.1; -
DR EMBL: BC010151; AAH10151.1; -
DR EMBL: BC014653; AAH14653.1; -
DR EMBL: AB061820; BAB79458.1; -
DR PIR: S37197; S37197.
DR PIR: S39803; S39803.
DR SWISS-2DPAGE: P36578; HUMAN.
DR Gene: HGNC:10353; RPL4.
DR MIM: 180479; -
DR InterPro: IPR002136; Ribosomal_L4/L1E.
DR Pfam: PF00573; Ribosomal_L4.1
DR PROSITE: PS00939; RIBOSOMAL_L1E; 1.
KW Ribosomal protein.
FT DOMAIN 364 427
FT CONFIDENT 3 3 LYS-RICH.
FT CONFIDENT 36 36 I -> V (IN REF. 1).
FT CONFIDENT 63 63 S -> R (IN REF. 1).
FT CONFIDENT 147 147 V -> F (IN REF. 1).
FT CONFIDENT 201 201 MISSING (IN REF. 1).
FT SEQUENCE 427 AA; 47697 MW; 4785ED31699CD792 CRC64;
Query Match 4.1%; Score 7; DB 1; Length 427;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 AAAAAL 14
DB 355 AAAAAL 361
RESULT 40
TUD4_AGRV1 STANDARD: PRT: 438 AA.
ID TUD4_AGRV1

```

AC 044472;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative hydroxypyruvate reductase (EC 1.1.1.81).
GN TTUJ.
OS Agrobacterium vitis (Rhizobium vitis).
OC Plasmid pTRAB4.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=373;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB4;
RX MEDLINE=96062236; PubMed=7592429;
RA Crouzet P., Otten L.;
RT "Sequence and mutational analysis of a tartrate utilization operon
   from Agrobacterium vitis."
RL J. Bacteriol. 177:6518-6526(1995).
CC -1- FUNCTION: DEGRADATES AN UNIDENTIFIED TOXIC PRODUCT FROM THE FIRST
CC -1- STEP OF TARTRATE DEGRADATION.
CC -1- CATALYTIC ACTIVITY: D-glycerate + NAD(P)(+) = hydroxypyruvate +
CC -1- NAD(P)H.
CC -1- PATHWAY: Tartrate degradation; second step.
CC -1- INDUCTION: By TARTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U25634; AAA68699.1;
DR InterPro: IPR005346; UPP0125.
DR Pfam: PF03658; UPP0125; 1
DR Plasmid: Oxidoreductase; NADP.
KW SEQUENCE 438 AA; 44730 MW; 38C75BAFFB0636B CRC64;
SQ

```

Query Match

Best Local Similarity	4.18;	Score 7;	DB 1;	Length 438;
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

OY 8 AAAAAL 14
 |||||
 Db 276 AAAAAL 282

Search completed: July 6, 2003, 14:20:46
 Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:19:30 ; Search time 32 Seconds
(without alignments)
1107.503 Million cell updates/sec

Title: US-09-674-779B-2

Perfect score: 172

Sequence: 1 MMLHIQIAAAALSVLTFM.....IATGTNEANSQNRRAELSY 172

Scoring table: OLIGO

Searched: 671580 seqs, 206047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database:

SPTRMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	5.8	186	16 Q9PC85	Q9PC85 xylella fas
2	9	5.2	998	10 Q94EA6	Q94EA6 oryza sativ
3	9	5.2	2168	5 Q9VOM0	Q9VOM0 drosophila
4	9	5.2	2176	5 Q46112	Q46112 drosophila
5	8	4.7	143	16 Q9X0G9	Q9X0G9 thermocoga
6	8	4.7	249	10 Q8S1R2	Q8S1R2 oryza sativ
7	8	4.7	258	5 Q9B1U7	Q9B1U7 drosophila
8	8	4.7	268	2 Q93NP5	Q93NP5 escherichia
9	8	4.7	280	16 Q930C4	Q930C4 rhizobium m
10	8	4.7	306	5 Q9VVI8	Q9VVI8 rhizobium m
11	8	4.7	316	16 Q986T0	Q986T0 rhizobium 1
12	8	4.7	324	5 Q9NGI8	Q9NGI8 drosophila
13	8	4.7	324	5 Q9NGI8	Q9NGI8 drosophila
14	8	4.7	324	5 Q9NGI7	Q9NGI7 drosophila
15	8	4.7	324	5 Q9NGB4	Q9NGB4 drosophila
16	8	4.7	324	5 Q9N6K3	Q9N6K3 drosophila

17	8	4.7	324	5 Q9N6K2	Q9N6K2 drosophila
18	8	4.7	381	3 Q8X1X2	Q8X1X2 aspergillus
19	8	4.7	452	16 Q9J5Z1	Q9J5Z1 neisseria m
20	8	4.7	504	16 Q98IK5	Q98IK5 rhizobium 1
21	8	4.7	526	16 Q9HTN7	Q9HTN7 pseudomonas
22	8	4.7	541	10 Q9S7V5	Q9S7V5 arabidopsis
23	8	4.7	545	5 Q9S5L1	Q9S5L1 drosophila
24	8	4.7	547	2 Q9XAX4	Q9XAX4 pseudomonas
25	8	4.7	570	10 Q9M7S3	Q9M7S3 lolium pere
26	8	4.7	587	2 Q8VQV7	Q8VQV7 burkholderi
27	8	4.7	587	5 Q02427	Q02427 ciona intes
28	8	4.7	624	5 Q9W2T0	Q9W2T0 drosophila
29	8	4.7	637	5 Q9W0E6	Q9W0E6 drosophila
30	8	4.7	637	5 Q8T6B9	Q8T6B9 drosophila
31	8	4.7	644	10 Q9PRV3	Q9PRV3 oryza sativ
32	8	4.7	644	10 Q9PRV2	Q9PRV2 oryza sativ
33	8	4.7	651	17 Q9HKB4	Q9HKB4 thermoplasm
34	8	4.7	662	5 Q9V9F7	Q9V9F7 drosophila
35	8	4.7	783	11 Q9QY56	Q9QY56 mus musculu
36	8	4.7	824	5 Q960R1	Q960R1 drosophila
37	8	4.7	825	5 Q9EMZ8	Q9EMZ8 drosophila
38	8	4.7	975	5 Q962D2	Q962D2 drosophila
39	8	4.7	1015	5 Q9BZW9	Q9BZW9 drosophila
40	8	4.7	1164	5 Q9Y161	Q9Y161 drosophila
41	8	4.7	1321	5 Q96ZD1	Q96ZD1 drosophila
42	8	4.7	1323	5 Q9NMX4	Q9NMX4 drosophila
43	8	4.7	1376	5 Q9BMZ9	Q9BMZ9 drosophila
44	8	4.7	1412	5 Q9VKJ1	Q9VKJ1 drosophila
45	8	4.7	1976	2 Q9MXI8	Q9MXI8 bacillus sp
46	8	4.7	2090	5 Q9WZT1	Q9WZT1 drosophila
47	8	4.7	2703	5 Q9VEG7	Q9VEG7 drosophila
48	8	4.7	2715	5 Q9VEG7	Q9VEG7 drosophila
49	8	4.1	38	10 Q9S9D9	Q9S9D9 nicotiana t
50	8	4.1	71	12 Q91LE1	Q91LE1 white spot
51	8	4.1	78	5 Q9BMP6	Q9BMP6 culicoides
52	8	4.1	82	13 Q9YH51	Q9YH51 pseudopleur
53	8	4.1	82	13 Q99013	Q99013 pseudopleur
54	8	4.1	96	10 Q8M327	Q8M327 oryza sativ
55	8	4.1	105	10 Q943N0	Q943N0 oryza sativ
56	8	4.1	110	10 Q942Y1	Q942Y1 oryza sativ
57	8	4.1	114	10 Q94E39	Q94E39 oryza sativ
58	8	4.1	122	15 Q9Q1V4	Q9Q1V4 human immun
59	8	4.1	132	2 Q53557	Q53557 salmonella
60	8	4.1	142	10 Q9SEW0	Q9SEW0 lilium long
61	8	4.1	147	5 Q95S08	Q95S08 drosophila
62	8	4.1	149	11 Q94102	Q94102 oryza sativ
63	8	4.1	149	11 Q9D7P2	Q9D7P2 mus musculu
64	8	4.1	150	10 Q9SA42	Q9SA42 arabidopsis
65	8	4.1	158	5 Q9Y195	Q9Y195 drosophila
66	8	4.1	158	11 Q9CTR4	Q9CTR4 mus musculu
67	8	4.1	163	10 Q94GR5	Q94GR5 oryza sativ
68	8	4.1	163	16 Q9KZF6	Q9KZF6 streptomyce
69	8	4.1	165	16 Q07218	Q07218 mycobacteri
70	8	4.1	168	2 Q51489	Q51489 pseudomonas
71	8	4.1	168	5 Q9NEE1	Q9NEE1 leishmania
72	8	4.1	168	16 Q914Z4	Q914Z4 pseudomonas
73	8	4.1	168	16 Q98P85	Q98P85 rhizobium 1
74	8	4.1	170	16 Q4662	Q4662 brucella ab
75	8	4.1	170	16 Q8YMX7	Q8YMX7 anabaena sp
76	8	4.1	171	5 Q9VP11	Q9VP11 drosophila
77	8	4.1	172	16 Q8Y1F4	Q8Y1F4 ralsstonia s
78	8	4.1	173	11 Q922F9	Q922F9 mus musculu
79	8	4.1	174	10 Q9LMT2	Q9LMT2 oryza sativ
80	8	4.1	174	16 Q8XPF6	Q8XPF6 salmonella
81	8	4.1	174	17 Q9YBA7	Q9YBA7 atropyrum p
82	8	4.1	176	10 Q9SMY4	Q9SMY4 arabidopsis
83	8	4.1	176	16 Q926C3	Q926C3 rhizobium m
84	8	4.1	177	16 Q8U9U5	Q8U9U5 agrobacteri
85	8	4.1	179	16 Q9RTG4	Q9RTG4 leishmania
86	8	4.1	180	5 Q9NMR9	Q9NMR9 leishmania
87	8	4.1	184	12 Q919S1	Q919S1 white spot
88	8	4.1	185	16 Q9PEZ5	Q9PEZ5 xylella fas
89	8	4.1	186	2 Q30783	Q30783 chlamydophi

```

90 7 4.1 188 16 084605 084605 chlamydia t
91 7 4.1 192 16 0927C5 0927C5 chlamydia p
92 7 4.1 196 5 09N7X4 09N7X4 leishmania
93 7 4.1 197 2 09X6J7 09X6J7 vibrio chol
94 7 4.1 198 10 09FTW8 09FTW8 oryza sativ
95 7 4.1 199 2 08VOT5 08VOT5 vibrio chol
96 7 4.1 199 8 08VOT4 08VOT4 vibrio chol
97 7 4.1 199 8 002667 002667 podospora a
98 7 4.1 200 10 09LN12 09LN12 arabidopsis
99 7 4.1 202 16 09PJE0 09PJE0 chlamydia m
100 7 4.1 206 10 0946V3 0946V3 zea mays (m

```

ALIGNMENTS

```

RESULT 1
ID 09PC85 PRELIMINARY; PRT; 186 AA.
AC 09PC85;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Outer membrane protein Pe precursor.
GN XF1896.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
NCBI_TaxID=2371;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Boye J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., Eldorrry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Klieger J.E., Kurama E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nant A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawaaski H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terezi M.F., Trufi D., Tsai S.M., Tshahko M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zaitz M., Zeldanis J., Zetubal J.C.;
RT "the genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
DR EMBL; AE004009; AAF84702.1; -.
DR InterPro; IPR001145; Bac_OmpA.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR ProDom; PD000930; Bac_OmpA; 1.
KW Complete proteome.
SQ
SEQUENCE 186 AA; 20431 MW; C8DD2FE233DB0C92 CRC64;

```

```

Query Match 5.8%; Score 10; DB 16; Length 186;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 108 TDERGSRVYN 117
Db 119 TDERGSRVYN 128

```

```

RESULT 2
ID 094EA6 PRELIMINARY; PRT; 998 AA.
AC 094EA6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE P0435H01.25 protein.
GN P0435H01.25.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
NCBI_TaxID=4530;
RX NCB1_TaxID=4530;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nlpnponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0435H01."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003142; BAB63529.1; -.
DR InterPro; IPR000862; RPOdomain.
SQ
SEQUENCE 998 AA; 108459 MW; 0826071700585730 CRC64;

```

```

Query Match 5.28; Score 9; DB 10; Length 998;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 8 AAAAALSV 16
Db 170 AAAAALSV 178

```

```

RESULT 3
ID 09VOM0 PRELIMINARY; PRT; 2168 AA.
AC 09VOM0;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE TOC protein.
GN TOC OR C93660.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dunin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C.J., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```


RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegawa C.,
 RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E.S., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Vaytskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003581; AAF51145.1; -
 DR FLYbase: FBgn0015600; loc.
 DR InterPro: IPR000169; SHPOT_acsite.
 DR PROSITE: PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
 SQ SEQUENCE 2168 AA; 234786 MW; F26344DB2E1CAF4D CRC64;

Query Match 5.2%; Score 9; DB 5; Length 2168;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 16
 Db 979 AAAAALSV 987

RESULT 4
 ID 046112 PRELIMINARY; PRT; 2176 AA.
 AC 046112;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE TOUCAN protein.
 GN TOC OR CG9660.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ADULT;
 RA MEDLINE=98090047; PubMed=9362455;
 RA Grammont M., Dastugue B., Couderc J.L.;
 RT "The *Drosophila* toucan (toc) gene is required in germline cells for
 RL development 124:4917-4926(1997).
 DR EMBL: Y14157; CAA74574.1; -
 DR FLYbase: FBgn0015600; loc.
 DR InterPro: IPR000169; SHPOT_acsite.
 DR PROSITE: PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
 SQ SEQUENCE 2176 AA; 235405 MW; 35ABDE00B49EFC7 CRC64;

Query Match 5.2%; Score 9; DB 5; Length 2176;
 Best Local Similarity 100.0%; Pred. No. 7.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 16
 Db 973 AAAAALSV 981

RESULT 5
 ID 09X0G9 PRELIMINARY; PRT; 143 AA.
 AC 09X0G9;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sugar-phosphate isomerase.
 GN TM1080.
 OS *Thermotoga maritima*.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571.
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RL genome sequence of *Thermotoga maritima*,"
 DR Nature 399:323-329(1999).
 DR EMBL: AE001768; AAD36157.1; -
 DR TIGR: TM1080; -
 DR InterPro: IPR003500; Rib/Gal_isomerase.
 DR Pfam: PF02502; LacAB_rfb; 1.
 DR TIGRFAMs: TIGR00689; rfb_LacA_LacB; 1.
 KM Isomerase; Complete proteome.
 SQ SEQUENCE 143 AA; 13867 MW; 3E17D72A10FCCBD CRC64;

Query Match 4.7%; Score 8; DB 16; Length 143;
 Best Local Similarity 100.0%; Pred. No. 7.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 NYLKGKI 137
 Db 20 NYLKGKI 27

RESULT 6
 ID 08S1R2 PRELIMINARY; PRT; 249 AA.
 AC 08S1R2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative UDP-glucuronic acid decarboxylase.
 GN P0506H12.31.
 OS *Oryza sativa* (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; *Oryza*.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC
 RL clone: P0506B12,"
 DR Submitted (FEB-2001) to the EMBL/GenBank/DBD databases.
 DR EMBL: AP003271; BAB89759.1; -
 SQ SEQUENCE 249 AA; 26592 MW; 7954F613B241E81F CRC64;

Query Match 4.7%; Score 8; DB 10; Length 249;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 SKLPSIVY 71
 Db 11111111

Db 25 SKIPLSVY 32

RESULT 7

ID 09BIJ7 PRELIMINARY; PRT; 258 AA.

AC 09BIJ7; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE DNA-binding transcriptional repressor giant (Fragment).

GN GR.
 OS Drosophila hydei (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7224;

RP SEQUENCE FROM N.A.
 RA Thomas J., Arnosti D.N.,
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF356543; AAK28631.1; -
 DR FlyBase; FBgn0044278; Dhyd\Nt.

FT DNA-binding. 1
 FT NON_TER 258 258
 FT SEQUENCE 258 AA; 26948 MW; 09C0CA545CE455EA CRC64;

Query Match 4.7%; Score 8; DB 5; Length 258;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALS 15
 Db 31 AAAAALS 38

RESULT 8

ID 093NP5 PRELIMINARY; PRT; 268 AA.

AC 093NP5; 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Mbwc.

GN Escherichia coli.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;

RP SEQUENCE FROM N.A.
 RA MEDLINE-21297197; PubMed-11404020;
 RA Wang L., Briggs C.E., Rothmund D., Fratamico P., Luchansky J.B.,
 RA Reeves P.R.;
 RT "Sequence of the E. coli O104 antigen gene cluster and identification of O104 specific genes.";
 RL Gene 270:231-236(2001).

DR EMBL; AF361371; AAK64375.1; -
 DR InterPro: IPR001173; Glycos.transf_2.
 DR Pfam: PF00535; Glycos.transf_2; 1.
 DR SEQUENCE 268 AA; 31176 MW; C1EF30C136A668F7 CRC64;

Query Match 4.7%; Score 8; DB 2; Length 268;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 TGVYITGV 49
 Db 239 TGVYITGV 246

RESULT 9

O930C4 ID 0930C4 PRELIMINARY; PRT; 280 AA.

AC 0930C4; 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein RA0274.

GN RA0274 OR SMA0520.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;

RP SEQUENCE FROM N.A.
 RA SPRAIN-1021;
 RA MEDLINE-21396509; PubMed-11481432;

RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
 RA Gurjal M., Hong A., Hulzar L., Hyman R.W., Kahn D., Kahn M.L.,
 RA Kallman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 RT "Nucleotide sequence and predicted functions of the entire
 Sinorhizobium meliloti pSym megaplasmid."

RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 DR EMBL; AE007220; AAK64932.1; -
 KW Plasmid; Hypothetical protein; Complete proteome.
 SEQUENCE 280 AA; 30615 MW; FF533F65700FC352 CRC64;

Query Match 4.7%; Score 8; DB 16; Length 280;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALS 15
 Db 121 AAAAALS 128

RESULT 10

ID 09VVI8 PRELIMINARY; PRT; 306 AA.

AC 09VVI8; 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE C66273 protein.

GN C66273.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.
 RA STRAIN-BERKELEY;
 RA MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gccayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek G., Nelson C.R., Miklos G.L.G.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Andrews-Pfannkoch C., Baldwin D.,
 RA Abell J.F., Agbayani A., An H.-J., Bayraktaroglu L., Beasley E.M.,
 RA Balow R.M., Basu A., Baxendale J., Bhandari D., Bolintinas S.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotler P.,
 RA Borrova D., Botchan M.R., Bouck J., Brackner J., Cantor A., Chandra I.,
 RA Burdick K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ikegawa C.,
 RA Jaleli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., Moberg D.,
 RA Mount S.M., Moy M., Murphy B., Murphy C., Morris J., Moshrefi A.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paul V., Reese M.G.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03523; AAF49323.1; -
 DR FlyBase: FBgn0036739; CG6273.
 SO SEQUENCE 306 AA; 30932 MW; 8C26CE0BEEDDADA3 CRC64;

Query Match 4.7%; Score 8; DB 5; Length 306;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAAAAAAL 14
 Db 195 TAAAAAAL 202

RESULT 11
 O986T0 PRELIMINARY; PRT; 316 AA.
 AC O986T0:
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein mlr7223.
 GN MLR7223.
 OS Rhizobium loti (Mesorhizobium loti).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Matsumoto A., Iida S., Ishikawa A., Kawasumi M., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimoto S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003011; BAB5373.1; -
 DR InterPro: IPR002173; Pfam.
 DR Pfam: PF00294; Pfam: 1.
 DR PROSITE: PS00584; PFAM_KINASES_2; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 316 AA; 32961 MW; 258AE2AVBE89DFID CRC64;

Query Match 4.7%; Score 8; DB 16; Length 316;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 Db 269 AAAAAAALS 276

RESULT 12
 O9NGI9 PRELIMINARY; PRT; 324 AA.
 AC O9NGI9:
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Eye1d (Fragment).
 GN OSA.
 OS Drosophila simulans (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7240;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SIM3;
 RX MEDLINE=20283933; PubMed=10823947;
 RA Begun D.J., Whitely P.;
 RT "Reduced X-linked nucleotide polymorphism in *Drosophila simulans*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
 DR EMBL: AF252666; AAF68048.1; -
 DR FlyBase: FBgn0041660; DsimVosa.
 FT NON_TER 1 324 1
 FT NON_TER 324 324 1
 SO SEQUENCE 324 AA; 32437 MW; 446774235A5071F0 CRC64;

Query Match 4.7%; Score 8; DB 5; Length 324;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 Db 269 AAAAAAALS 276

RESULT 13
 O9NGI8 PRELIMINARY; PRT; 324 AA.
 AC O9NGI8:
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Eye1d (Fragment).
 GN OSA.
 OS Drosophila simulans (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7240;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SIM4;
 RX MEDLINE=20283933; PubMed=10823947;
 RA Begun D.J., Whitely P.;
 RT "Reduced X-linked nucleotide polymorphism in *Drosophila simulans*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
 DR EMBL: AF252666; AAF68049.1; -
 DR FlyBase: FBgn0041660; DsimVosa.
 FT NON_TER 1 324 1
 FT NON_TER 324 324 1
 SO SEQUENCE 324 AA; 32373 MW; 5BCC6A8C4546A230 CRC64;

Query Match 4.7%; Score 8; DB 5; Length 324;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 Db 269 AAAAAAALS 276

RESULT 14
ID 09NG17 PRELIMINARY; PRT: 324 AA.
AC 09NG17;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EyeIid (Fragment).
GN OSA.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SIM5;
RX MEDLINE-20283933; PubMed=10823947;
RA Begun D.J., Whitley P.;
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF252668; AAF68050.1; -;
DR FlyBase; FBgn0041660; Dsim\osa.
FT NON_TER 1 324
SQ SEQUENCE 324 AA; 32399 MW; 5ECC6A8C4556B331 CRC64;
Query Match 4.7%; Score 8; DB 5; Length 324;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 AAAAALS 15
DB 269 AAAAALS 276

RESULT 15
ID 09NGB4 PRELIMINARY; PRT: 324 AA.
AC 09NGB4;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EyeIid (Fragment).
GN OSA.
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-20283933; PubMed=10823947;
RA Begun D.J., Whitley P.;
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF255314; AAF68611.1; -;
DR FlyBase; FBgn0041640; Dyak\osa.
FT NON_TER 1 324
SQ SEQUENCE 324 AA; 32644 MW; 65BA271ED4DE03D CRC64;
Query Match 4.7%; Score 8; DB 5; Length 324;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 AAAAALS 15
DB 269 AAAAALS 276

09NGK3 PRELIMINARY; PRT: 324 AA.
AC 09NGK3;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EyeIid (Fragment).
GN OSA.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SIM7, AND SIM2;
RX MEDLINE-20283933; PubMed=10823947;
RA Begun D.J., Whitley P.;
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF252670; AAF68052.1; -;
DR EMBL; AF252665; AAF68047.1; -;
DR FlyBase; FBgn0041660; Dsim\osa.
FT NON_TER 1 324
SQ SEQUENCE 324 AA; 32403 MW; 4461ADE6E896B334 CRC64;
Query Match 4.7%; Score 8; DB 5; Length 324;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 AAAAALS 15
DB 269 AAAAALS 276

RESULT 17
ID 09NGK2 PRELIMINARY; PRT: 324 AA.
AC 09NGK2;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EyeIid (Fragment).
GN OSA.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SIM8, AND SIM6;
RX MEDLINE-20283933; PubMed=10823947;
RA Begun D.J., Whitley P.;
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF252671; AAF68053.1; -;
DR EMBL; AF252669; AAF68051.1; -;
DR FlyBase; FBgn0041660; Dsim\osa.
FT NON_TER 1 324
SQ SEQUENCE 324 AA; 32407 MW; 44676F2E5A5071F0 CRC64;
Query Match 4.7%; Score 8; DB 5; Length 324;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 AAAAALS 15
DB 269 AAAAALS 276

```

RESULT 18
ID 08X1X2 PRELIMINARY; PRT; 381 AA.
AC 08X1X2;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
PE Pectin lyase A.
GN PECTA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitospotic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KBN616;
RA Kitamoto N.;
RT "Two pectin lyase genes (pecta and pectb) from Aspergillus oryzae
RT KBN616: their sequence analyses and overexpression, and
RT characterization of the gene products."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB029322; BAB82467.1;
DR InterPro: IPR002022; Amb_allergen.
DR Pfam: PF00544; pec_lyase; 1.
KW Lyase.
SQ SEQUENCE 381 AA; 39982 MW; BA6FC139D31E10F7 CRC64;

Query Match
Best Local Similarity 4.7%; Score 8; DB 16; Length 381;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAAAL 14
DB 7 IAAAAAAL 14

RESULT 19
ID 09J5Z1 PRELIMINARY; PRT; 452 AA.
AC 09J5Z1;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE UDP-MurNAC-pentapeptide synthetase (EC 6.3.2.15).
GN MURF OR NMA2068.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=63699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SPROTYPE 4A.
RX MEDLINE=20222556; PubMed=10761919;
RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jørgensen K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajendram M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
DR EMBL; AL162758; CAB85286.1;
DR HSSP; P11880; 16G4.
DR InterPro: IPR000713; Mur_lyase.
DR InterPro: IPR004101; Mur_lyase_C.
DR Pfam: PF01225; Mur_lyase; 1.
DR Pfam: PF02875; Mur_lyase_C; 1.
DR TIGRfams; TIGR01143; murf; 1.
KW Ligase; Complete proteome.
SQ SEQUENCE 452 AA; 48112 MW; 9903D4DD1B64C49A CRC64;

Query Match
Best Local Similarity 4.7%; Score 8; DB 16; Length 452;
Matches 100.0%; Pred. No. 20;

RESULT 20
ID 0981K5 PRELIMINARY; PRT; 504 AA.
AC 0981K5;
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical protein mlr2359.
GN MLR2359.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno M.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002999; BAB49511.1;
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA; 1.
DR ProDom: PD000930; Bac_OmpA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 504 AA; 53754 MW; ABDCC1D5759DD0DB CRC64;

Query Match
Best Local Similarity 4.7%; Score 8; DB 16; Length 504;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAAAL 14
DB 282 IAAAAAAL 289

RESULT 21
ID 09HTN7 PRELIMINARY; PRT; 526 AA.
AC 09HTN7;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Probable binding protein component of ABC dipeptide transporter.
GN PA5317.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen."

```

RL Nature 406:959-964(2000).
 DR EMBL: AE004944; AAG08702.1; --
 DR HSSP: P23847; 1DPE.
 DR InterPro: IPR000914; SBP_bac_5.
 DR Pfam: PF00496; SBP_bac_5; 1.
 DR Complete proteome.
 SW SEQUENCE 526 AA; 57880 MW; EE3F4B3A0914B24B CRC64;

Query Match 4.7%; Score 8; DB 16; Length 526;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AAAAALSV 16
 Db 19 AAAAALSV 26

RESULT 22

ID 0957V5 PRELIMINARY; PRT; 541 AA.
 AC 0957V5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE T16011.4 protein (AF360900/16011_4).
 GN T16011.4 OR MZB10.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

RA SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronning C.M., Koo H., Fujii C.Y., Uterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III BAC T16011 genomic sequence.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RA SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronning C.M., Koo H., Fujii C.Y., Uterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III P1 MZB10 genomic sequence.";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RA SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Chen R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bower L., Carinici P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.T., Sakurai T.,
 RA Satou M., Seki Y., Sultwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;

RT "Arabidopsis cDNA clones.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC010871; AAF07827.1; --
 DR EMBL: AC009326; AAD56316.1; --
 DR EMBL: AF428276; AAL16108.1; --
 SW SEQUENCE 541 AA; 56813 MW; 3AEF51AD4EAD6E CRC64;

Query Match 4.7%; Score 8; DB 10; Length 541;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 15
 Db 43 AAAAALSV 50

RESULT 23

ID 095S11 PRELIMINARY; PRT; 545 AA.
 AC 095S11;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE H102140P.
 GN PUBSF OR CG12085.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;

RA SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Broksstein P., Hong L., Agapayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF061006; AAL28554.1; --
 DR Flybase: FBgn028577; pubsf.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 3.
 DR PROSITE: PS50102; RRM; 3.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_2.
 SW SEQUENCE 545 AA; 57989 MW; E75E378FD07E24E CRC64;

Query Match 4.7%; Score 8; DB 5; Length 545;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 15
 Db 369 AAAAALSV 376

RESULT 24

ID 09X4X4 PRELIMINARY; PRT; 547 AA.
 AC 09X4X4;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE DITE.
 GN DITE.
 OS Pseudomonas abietaniphila.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=89065;

RA SEQUENCE FROM N.A.
 RC STRAIN=BKME-9;
 RA Mohn W.W., Wilson A.E., Bicho P., Moore E.R.B.;
 RT "Physiological and Phylogenetic Diversity of Bacteria Growing on Resin
 Acids.";
 RL Syst. Appl. Microbiol. 0:0-0(1999).
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=BKME-9;
 RX MEDLINE=99235742; PubMed=10217753;

RA Martin V.J., Mohn W.W.;
 RT "A novel aromatic-ring-hydroxylating dioxygenase from the diterpenoid-
 degrading bacterium Pseudomonas abietaniphila BKME-9.";
 RL J. Bacteriol. 181:2675-2682(1999).
 DR EMBL: AF119621; AAD21067.1; --
 SW SEQUENCE 547 AA; 58500 MW; 49A627265D7C1A74 CRC64;

Query Match 4.7%; Score 8; DB 2; Length 547;
 Best Local Similarity 100.0%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAALSVL 17
 |||||
 Db 107 AAAALSVL 114

RESULT 25

O9M7S3 PRELIMINARY; PRT; 570 AA.

AC 09M7S3; 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 OS 4-comarate-coa ligase 4CL1 (EC 6.2.1.12).
 OC Lollum perenne (Perennial ryegrass).
 GN Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Poace; Lollum.
 OX NCBI_TaxID=4522;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. ELLET;
 RA Heath R.L., Huxley H., Spangenberg G.;
 RT "Isolation of three 4-comarate-coa ligase cDNA homologues from
 Lollum perenne".
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF032221; AAF3732.1; -.
 DR HSSP; P08659; 1LCI.
 DR InterPro: IPR000873; AMP-Bind.
 DR Pfam; PF00501; AMP-binding; 1.
 DR PROSITE; PS00455; AMP_BINDING; 1.
 KW Ligase.
 SQ SEQUENCE 570 AA; 60293 MW; 8194E1B939D2925 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 8; DB 10; Length 570;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 QIAAAAAA 13
 |||||
 Db 13 QIAAAAAA 20

RESULT 26

O8VOP7 PRELIMINARY; PRT; 587 AA.

AC 08VOP7; 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 GN Flagellar MS ring protein.
 OS Burkholderia cepacia (Pseudomonas cepacia).
 OC Bacteria; Proteobacteria; Beta subdivision; Burkholderia group;
 OC Burkholderia.
 OX NCBI_TaxID=292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J2315;
 RA Tomich M., Herfst C.A., Golden J.W., Mohr C.D.;
 RT "Role of Flagella in Burkholderia cepacia Host Cell Invasion."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF453480; AAL65160.1; -.
 DR InterPro: IPR000067; FlgMing_FLIIF.
 DR InterPro: IPR002920; YscJ_FLIIF.
 DR Pfam; PF01514; YscJ_FLIIF; 1.
 DR PRINTS; PR01009; FLGMRINGFLIF.
 DR TIGRfams; TIGR00206; flif; 1.
 SQ SEQUENCE 587 AA; 62991 MW; 4B142D53E065CB91 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 8; DB 2; Length 587;
 Matches 100.0%; Pred. No. 25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 IAAAAAAL 14
 |||||
 Db 491 IAAAAAAL 498

RESULT 27

O02427 PRELIMINARY; PRT; 587 AA.

AC 002427; 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 GN Forkhead homolog.
 GN CI-FKH.
 OS Clona intestinalis.
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 OC Phlebobranchia; Clonidae; Clona.
 OX NCBI_TaxID=7719;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Corbo J.C., Erives A., Di Gregorio A., Chang A., Levine M.;
 RT "Conserved patterning of the vertebrate neural tube is conserved in
 a protochordate."
 RL Development 0:0-0(1997).
 DR EMBL; AF002988; AAB61227.1; -.
 DR HSSP; Q63245; 2HFH.
 DR InterPro: IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS00039; FORK_HEAD_3; 1.
 SQ SEQUENCE 587 AA; 64092 MW; 08051624221A9D00 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 8; DB 5; Length 587;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 QIAAAAAA 13
 |||||
 Db 80 QIAAAAAA 87

RESULT 28

O9W2T0 PRELIMINARY; PRT; 624 AA.

AC 09W2T0; 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 GN CG15295 protein.
 GN CG15295.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelhiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA April J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kaulush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Jatali M., Kaulush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Mekulov G., Mlshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Merkulov G., Mlshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
 RA Palazzolo M., Pittman G.S., Pan S., Saunders R.D.C., Scheeler F., Shen H.,
 RA Reinert K., Remington K., Saunders R.D.C., Skupski M.P., Smith T.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spleier E., Spradling A.C., Stappleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).
 DR EMBL: AF003451; AAF46609.1; -
 DR Flybase: FBgn0030220; CG15995.
 DR FLYBASE: FBgn0030220; CG15995.
 SQ SEQUENCE 624 AA: 65843 MW: 8960041374AFD6C1 CRC64;

Query Match 4.7%; Score 8; DB 5; Length 624;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||||
 Db 158 AAAAAAALS 165

RESULT 29
 O9W0E6 PRELIMINARY; PRT: 637 AA.
 ID O9W0E6; Q9W0E7; Q9U696; 13, Created)
 AC Q9W0E6; Q9W0E7; Q9U696; 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Poly-U-binding-splicing-factor protein (CG12085 protein).
 GN PUSF OR POLY-U-BINDING-SPlicing-FACTOR OR CG12085.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu U., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kaulush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Mekulov G., Mlshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Merkulov G., Mlshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
 RA Palazzolo M., Pittman G.S., Pan S., Saunders R.D.C., Scheeler F., Shen H.,
 RA Reinert K., Remington K., Saunders R.D.C., Skupski M.P., Smith T.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spleier E., Spradling A.C., Stappleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RA Page-McCaw P.S., Anonlidivman K., Sharp P.A.;
 RT "P6060: A pyrimidine-tract binding splicing factor and U2AF55
 homolog.";
 RT RNA 0:0-0(1999).
 RT EMBL: AE003471; AAF47501.1; -
 DR EMBL: AE003471; AAF47501.1; -
 DR EMBL: AE003471; AAF47501.1; -
 DR HSP: P26368; 202E.
 DR Flybase: FBgn0028577; pubst.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 3.
 DR SMART: SM00360; RRM; 3.
 DR PROSITE: PS0102; RRM; 3.
 DR PROSITE: PS0030; RRM_RNP_1; 2.
 DR PROSITE: PS0030; RRM_RNP_1; 2.
 SQ SEQUENCE 637 AA: 67939 MW: F0AD184D9E8F145F CRC64;

Query Match 4.7%; Score 8; DB 5; Length 637;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||||
 Db 461 AAAAAAALS 468

RESULT 30
 O8T6B9 PRELIMINARY; PRT: 637 AA.
 ID O8T6B9; Q8T6B9; 21, Created)
 AC O8T6B9; Q8T6B9; 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Half pInt.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Van Buskirk C., Schubach T.;
 RA "half pInt regulates alternative splice site selection in
 RT Drosophila.";
 RT Dev. Cell 0:0-0(2002).
 DR EMBL: AF479079; AAL86452.1; -

SO SEQUENCE 637 AA; 67895 MW; 00A81D4D9EEF1282 CRC64;
 Query Match 4.7%; Score 8; DB 5; Length 637;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||
 Db 461 AAAAAAALS 468

RESULT 31

O9FRV3 PRELIMINARY; PRT; 644 AA.
 AC O9FRV3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Acetolactate synthase.
 GN ALS.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eumariophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriarioidae; Oryzae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. KINMAZE;
 RA Shimizu T., Kato Y., Nakayama I., Nakayama K., Fukuda A., Tanaka Y.;
 RT "Isolation and Expression of acetolactate synthase genes from Oryza
 sativa.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
 DR EMBL: AB049823; BAB20812.1;
 DR InterPro: IPR004407; AcoIac_19.
 DR InterPro: IPR000399; TPP_enzyme.
 DR Pfam: PF02775; TPP_enzymes.C; 1.
 DR Pfam: PF02776; TPP_enzymes.N; 1.
 DR TIGRfams: TIGR00118; acoIac_19; 1.
 DR KEGG: TIGR00118; acoIac_19; 1.
 KW Flavoprotein; Lyase; Thiamine pyrophosphate.
 SO SEQUENCE 644 AA; 69392 MW; EDE26E5F8B89EC9 CRC64;

Query Match 4.7%; Score 8; DB 10; Length 644;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||
 Db 6 AAAAAAALS 13

RESULT 32

O9FRV2 PRELIMINARY; PRT; 644 AA.
 AC O9FRV2;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Acetolactate synthase.
 GN ALS.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eumariophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriarioidae; Oryzae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. KINMAZE;
 RA Shimizu T., Kato Y., Nakayama I., Nakayama K., Fukuda A., Tanaka Y.;
 RT "Isolation and Expression of acetolactate synthase genes from Oryza
 sativa.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

CC -1 SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
 DR EMBL: AB049823; BAB20813.1;
 DR InterPro: IPR004407; AcoIac_19.
 DR InterPro: IPR000399; TPP_enzyme.
 DR Pfam: PF02775; TPP_enzymes.C; 1.
 DR Pfam: PF02776; TPP_enzymes.N; 1.
 DR TIGRfams: TIGR00118; acoIac_19; 1.
 DR KEGG: TIGR00118; acoIac_19; 1.
 KW Flavoprotein; Lyase; Thiamine pyrophosphate.
 SO SEQUENCE 644 AA; 69345 MW; F873F121A9FD9F99 CRC64;

Query Match 4.7%; Score 8; DB 10; Length 644;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||
 Db 6 AAAAAAALS 13

RESULT 33

O9HKB4 PRELIMINARY; PRT; 651 AA.
 AC O9HKB4;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE Hypothetical protein Ta0687.
 GN TA0687.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Rupp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mepp H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermophilic scavenger Thermoplasma
 acidophilum.";
 RL Nature 407:508-513 (2000).
 DR EMBL: AL445065; CAC11825.1;
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 651 AA; 74410 MW; 3F9E9F1711E7B733 CRC64;

Query Match 4.7%; Score 8; DB 17; Length 651;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 LHIQTAA 10
 |||||
 Db 128 LHIQTAA 135

RESULT 34

O9V9F7 PRELIMINARY; PRT; 662 AA.
 AC O9V9F7;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CG10417 protein (LD27655P).
 GN CG10417.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;

Query Match 4.7%; Score 8; DB 5; Length 637;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||
 Db 461 AAAAAAALS 468

09BMZ8

ID 09BMZ8 PRELIMINARY; PRT; 825 AA.

AC 09BMZ8; 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

GN AF10.

OS ALHAMBRA OR CG1070.

OC Drosophila melanogaster (Fruit fly).

CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

RN NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.

RA Linder B., Jackie H.;

RT "The Drosophila homolog of AF10 is expressed in a striped pattern

RL throughout blastoderm."

DR Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF281146; AAK06386.1;

DR FlyBase; FBgn0037471; Alhambra.

SQ SEQUENCE 825 AA; 84483 MW; 9A2C7A795B9F43D6 CRC64;

Query Match

Best Local Similarity 4.7%; Score 8; DB 5; Length 825;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 8 AAAAAAALS 15

DB 601 AAAAAAALS 608

RESULT 38

ID 0962D2

AC 0962D2;

DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

GN ALHAMBRA OR CG1070.

OS Drosophila melanogaster (Fruit fly).

CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

RN NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.

RA MEDLINE=21095180; PubMed=11165485;

RA Bahri S.M., Chia W., Yang X.;

RT "The Drosophila homolog of human AF10/AF17 leukemia fusion genes

RT (Daf) encodes a zinc finger/leucine zipper nuclear protein regulated

RT growth."

RT Mech. Dev. 100:291-301(2001).

DR EMBL; AY036104; AAK63170.1;

DR FlyBase; FBgn0037471; Alhambra.

SQ SEQUENCE 975 AA; 98713 MW; 5BB70FB144935276 CRC64;

Query Match

Best Local Similarity 4.7%; Score 8; DB 5; Length 975;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 8 AAAAAAALS 15

DB 600 AAAAAAALS 607

RESULT 39

ID 08SZW9

AC 08SZW9; PRELIMINARY; PRT; 1015 AA.

DT 01-JUN-2002 (TREMBlrel. 21, Created)

ID 09YI61

AC 09YI61; PRELIMINARY; PRT; 1164 AA.

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)

GN ALHAMBRA OR CG1070.

OS Drosophila melanogaster (Fruit fly).

CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

RN NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.

RA STRAIN=BERKELEY;

RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amandalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA Sutton G.G., Mortman J.R., Richards S., Ashburner M., Henderson S.N.,

RA Brandon R.C., Rogers T.H.C., Blazek R.G., Zhang O., Chen L.X.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,

RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Broxstein P., Brotlier P.,

RA Burlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Dugas-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jatala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Modyary C., Morris J., Koshneff A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissendach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.F., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
DR EMBL: AF003672; AAF54065.1; -
DR FlyBase: FBgn0037471; Alhambra.
SQ SEQUENCE 1164 AA; 119469 MW; F49C6BC7394DBB6A CRC64;

Query Match 4.7%; Score 8; DB 5; Length 1164;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALS 15
|||||||
DB 940 AAAAALS 947

Search completed: July 6, 2003, 14:21:32
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:20:50 ; Search time 19 Seconds

(without alignments)
266.355 Million cell updates/sec

Title: US-09-674-779b-2

Perfect score: 172
Sequence: 1 MMLHIOIAAAALSLVTFM.....IAFTNEAMQNRRAELSY 172

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	4.1	9	US-08-159-339A-154	Sequence 154, App
2	7	4.1	13	PCT-US95-04121-52	Sequence 52, App
3	7	4.1	14	US-08-503-226B-42	Sequence 42, App
4	7	4.1	14	US-08-721-458B-42	Sequence 42, App
5	7	4.1	22	US-08-516-859A-103	Sequence 103, App
6	7	4.1	22	US-09-586-472-103	Sequence 103, App
7	7	4.1	22	US-09-528-706-103	Sequence 103, App
8	7	4.1	37	US-08-180-524-1	Sequence 1, App
9	7	4.1	37	US-08-180-524-8	Sequence 8, App
10	7	4.1	37	US-08-180-524-9	Sequence 9, App
11	7	4.1	37	US-08-975-166-1	Sequence 1, App
12	7	4.1	37	US-08-975-166-8	Sequence 8, App
13	7	4.1	37	US-08-975-166-9	Sequence 9, App
14	7	4.1	37	US-09-117-121-39	Sequence 39, App
15	7	4.1	37	US-09-117-121-42	Sequence 42, App
16	7	4.1	37	US-09-344-529-4	Sequence 4, App
17	7	4.1	38	US-07-814-220-1	Sequence 1, App
18	7	4.1	109	US-07-812-421-1	Sequence 1, App
19	7	4.1	109	US-09-325-932A-144	Sequence 144, App
20	7	4.1	203	US-08-233-609-3	Sequence 3, App
21	7	4.1	203	US-08-444-083-3	Sequence 3, App
22	7	4.1	203	US-08-286-304-3	Sequence 3, App
23	7	4.1	203	US-08-442-745-3	Sequence 3, App
24	7	4.1	203	US-08-443-129-3	Sequence 3, App
25	7	4.1	203	US-08-443-952-3	Sequence 3, App
26	7	4.1	203	US-08-443-130-3	Sequence 3, App
27	7	4.1	203	US-08-898-911-3	Sequence 3, App

28	7	4.1	203	PCT-US95-04467-3	Sequence 3, App
29	7	4.1	228	US-09-286-690-12	Sequence 12, App
30	7	4.1	309	US-08-508-761B-32	Sequence 32, App
31	7	4.1	325	US-08-107-676-30	Sequence 30, App
32	7	4.1	350	US-09-161-241-9	Sequence 9, App
33	7	4.1	482	US-07-792-885A-1	Sequence 1, App
34	7	4.1	482	US-08-142-439A-7	Sequence 7, App
35	7	4.1	482	US-08-869-477-7	Sequence 7, App
36	7	4.1	503	US-09-172-952-26	Sequence 26, App
37	7	4.1	503	US-09-382-256-10	Sequence 10, App
38	7	4.1	503	US-09-395-115-10	Sequence 10, App
39	7	4.1	503	US-08-436-285-10	Sequence 10, App
40	7	4.1	503	US-09-679-187-10	Sequence 10, App
41	7	4.1	674	US-08-893-852A-1	Sequence 1, App
42	7	4.1	760	US-08-195-152-2	Sequence 2, App
43	7	4.1	894	US-09-735-934A-4	Sequence 4, App
44	7	4.1	1084	US-09-237-725A-3	Sequence 3, App
45	7	4.1	1724	US-08-857-076-12	Sequence 12, App
46	7	4.1	7257	US-09-335-409-5	Sequence 5, App
47	7	4.1	7257	US-09-568-102-5	Sequence 5, App
48	7	4.1	7257	US-09-567-969-5	Sequence 5, App
49	7	4.1	7257	US-09-568-480-5	Sequence 5, App
50	7	4.1	7257	US-09-568-486-5	Sequence 5, App
51	7	4.1	7257	US-09-568-472-5	Sequence 5, App
52	7	4.1	7257	US-09-567-899-5	Sequence 5, App
53	6	3.5	6	US-08-837-305-1	Sequence 1, App
54	6	3.5	6	US-08-611-395-9	Sequence 9, App
55	6	3.5	7	US-09-122-126B-5	Sequence 5, App
56	6	3.5	8	US-08-963-168C-24	Sequence 24, App
57	6	3.5	8	US-08-963-168C-26	Sequence 26, App
58	6	3.5	9	US-08-425-069-12	Sequence 12, App
59	6	3.5	9	US-08-317-844B-12	Sequence 12, App
60	6	3.5	9	US-08-396-385-12	Sequence 12, App
61	6	3.5	9	US-08-159-339A-143	Sequence 143, App
62	6	3.5	9	US-08-159-339A-152	Sequence 152, App
63	6	3.5	9	US-08-159-339A-151	Sequence 151, App
64	6	3.5	9	US-08-159-339A-155	Sequence 155, App
65	6	3.5	9	US-08-159-339A-153	Sequence 153, App
66	6	3.5	9	US-08-159-339A-154	Sequence 154, App
67	6	3.5	9	US-08-159-339A-156	Sequence 156, App
68	6	3.5	9	US-08-159-339A-157	Sequence 157, App
69	6	3.5	9	US-08-159-339A-158	Sequence 158, App
70	6	3.5	9	US-08-159-339A-159	Sequence 159, App
71	6	3.5	9	US-08-159-339A-160	Sequence 160, App
72	6	3.5	9	US-08-159-339A-161	Sequence 161, App
73	6	3.5	9	US-08-159-339A-162	Sequence 162, App
74	6	3.5	9	US-08-159-339A-163	Sequence 163, App
75	6	3.5	9	US-08-159-339A-164	Sequence 164, App
76	6	3.5	9	US-08-159-339A-165	Sequence 165, App
77	6	3.5	9	US-08-159-339A-166	Sequence 166, App
78	6	3.5	9	US-08-159-339A-167	Sequence 167, App
79	6	3.5	9	US-08-159-339A-168	Sequence 168, App
80	6	3.5	9	US-08-159-339A-169	Sequence 169, App
81	6	3.5	9	US-08-159-339A-170	Sequence 170, App
82	6	3.5	9	US-08-159-339A-171	Sequence 171, App
83	6	3.5	9	US-09-287-221-12	Sequence 12, App
84	6	3.5	10	US-08-406-193-14	Sequence 14, App
85	6	3.5	10	US-08-463-863-9	Sequence 9, App
86	6	3.5	10	US-08-545-151-14	Sequence 14, App
87	6	3.5	10	US-08-458-887-9	Sequence 9, App
88	6	3.5	10	US-08-159-339A-149	Sequence 149, App
89	6	3.5	10	US-08-159-339A-150	Sequence 150, App
90	6	3.5	10	US-08-963-168C-25	Sequence 25, App
91	6	3.5	10	5169933-18	Sequence 18, App
92	6	3.5	10	5169933-19	Sequence 19, App
93	6	3.5	10	5169933-41	Sequence 41, App
94	6	3.5	10	US-07-814-220-5	Sequence 5, App
95	6	3.5	11	US-07-814-220-6	Sequence 6, App
96	6	3.5	11	US-07-814-220-8	Sequence 8, App
97	6	3.5	11	US-07-814-220-11	Sequence 11, App
98	6	3.5	11	US-07-812-421-5	Sequence 5, App
99	6	3.5	11		
100	6	3.5	11		

ALIGNMENTS

RESULT 1
US-08-159-339A-154
Sequence 154, Application US/08159339A
Patent No. 6037135
GENERAL INFORMATION:
APPLICANT: Kubo, Ralph T.
APPLICANT: Grey, Howard M.
APPLICANT: Settle, Alessandro
APPLICANT: Settle, Esteban
TITLE OF INVENTION: HLA Binding peptides and their
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 1254
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,339A
FILING DATE: 29-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,666
FILING DATE: 07-AUG-1992
APPLICATION NUMBER: US 08/027,746
FILING DATE: 05-MAR-1993
APPLICATION NUMBER: US 08/103,396
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 018623-005030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
TELEX:
INFORMATION FOR SEQ ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-159-339A-154

Query Match 4.1%; Score 7; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAAAAA 13
|||||
DB 2 TAAAAA 8

RESULT 2
PCT-US95-04121-52
Sequence 52, Application PC/TUS9504121
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Haptened peptides and Uses Thereof
NUMBER OF SEQUENCES: 62

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04121
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/222,206
FILING DATE: April 1, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Vanstone, Darlene A.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 079.2PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6010
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
PCT-US95-04121-52

Query Match 4.1%; Score 7; DB 5; Length 13;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAAAAA 13
|||||
DB 3 TAAAAA 9

RESULT 3
US-08-503-226B-42
Sequence 42, Application US/08503226B
Patent No. 5871945
GENERAL INFORMATION:
APPLICANT: Lockerdie, Robert Owen, et al.
TITLE OF INVENTION: Calcineurin Inhibitory Compounds and
TITLE OF INVENTION: Anchoring Protein
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/503,226B
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/404,731
FILING DATE: 15-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,227
FILING DATE: 23-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32861

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-6300
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-503-226B-42

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 5 AAAAAA 11

RESULT 4
US-08-721-458B-42
Sequence 42, Application US/08721458B
Patent No. 6107104
GENERAL INFORMATION:
APPLICANT: Lockebble, Robert Owen, et al.
TITLE OF INVENTION: Calcineurin Inhibitory Compounds
TITLE OF INVENTION: and Anchoring Protein
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721,458B
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/404,731
FILING DATE: 15-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,227
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/503,226
FILING DATE: 17-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/33276
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-6300
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-721-458B-42

Query Match
Best Local Similarity 4.1%; Score 7; DB 3; Length 14;
Matches 100.0%; Pred. No. 2.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 AAAAAA 14
Db 5 AAAAAA 11

RESULT 5
US-08-516-859A-103
Sequence 103, Application US/08516859A
Patent No. 6069231
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/516,859A
FILING DATE: 18-AUG-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-516-859A-103

Query Match
Best Local Similarity 4.1%; Score 7; DB 3; Length 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 6
US-09-586-472-103
Sequence 103, Application US/09586472
Patent No. 632335
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
TITLE OF INVENTION: Zinc Finger Proteins
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/09/586,472
FILING DATE: 01-Jun-2000
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/528,706
FILING DATE: 17-MAR-2000
APPLICATION NUMBER: US 08/516,859
FILING DATE: 18-AUG-1995
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 4130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-09-586-472-103

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 6 AAAAAA 12

RESULT 7
US-09-528-706-103
Sequence 103, Application US/09528706
Patent No. 6468985

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/528,706
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/516,859

FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-09-528-706-103

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 6 AAAAAA 12

RESULT 8
US-08-180-524-1
Sequence 1, Application US/08180524
Patent No. 5849537

GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Elzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudopleuronectes americanus
US-08-180-524-1

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 9
US-08-180-524-8
Sequence 8, Application US/08180524
Patent No. 5849537
GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Elzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thad Kryshak, Charles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-180-524-8

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 10
US-08-180-524-9
Sequence 9, Application US/08180524
Patent No. 5849537
GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Elzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thad Kryshak, Charles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-180-524-9

Query Match 4.1%; Score 7; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAL 14
| | | | |
Db 6 AAAAAL 12

RESULT 11
US-08-975-166-1
Sequence 1, Application US/08975166
Patent No. 5928877
GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kot, Edward
APPLICANT: Chilcove, Etzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,166
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Pseudopleuronectes americanus
US-08-975-166-1

Query Match 4.1%; Score 7; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAL 14
| | | | |
Db 6 AAAAAL 12

RESULT 12
US-08-975-166-8
Sequence 8, Application US/08975166
Patent No. 5928877
GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kot, Edward
APPLICANT: Chilcove, Etzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,166
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-975-166-8

Query Match 4.1%; Score 7; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 13

US-08-975-166-9
; Sequence 9, Application US/08975166

Patent No. 5928877

GENERAL INFORMATION:

APPLICANT: Tripp, Matthew

APPLICANT: Lusk, Lance

APPLICANT: Rhodes, Thomas

APPLICANT: Hulge, Nick

APPLICANT: Kol, Edward

APPLICANT: Chilcote, Elzer

APPLICANT: Barney, Michael C.

APPLICANT: Bower, Patricia A.

APPLICANT: Cronan, Charles L.

TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS

TITLE OF INVENTION: IN YEAST

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Thad Kryshak, Quarles & Brady

STREET: 411 East Wisconsin Avenue

CITY: Milwaukee

STATE: Wisconsin

COUNTRY: USA

ZIP: 53202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/975,166

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/180,524

FILING DATE:

APPLICATION NUMBER: US/07/917,216

FILING DATE:

APPLICATION NUMBER: US 07/486,333

FILING DATE: 28-FEB-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/409,217

FILING DATE: 19-SEP-1989

ATTORNEY/AGENT INFORMATION:

NAME: Kryshak, Thad

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 66-005-9234-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5707

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 37 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYDROPHETICAL: NO

ANTI-SENSE: NO

US-08-975-166-9

Query Match 4.1%; Score 7; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 14

US-09-117-121-39
; Sequence 39, Application US/09117121

Patent No. 6307020

GENERAL INFORMATION:

APPLICANT: Hew, Choy

APPLICANT: Gong, Zhiyuan

TITLE OF INVENTION: Intracellular Antifreeze Polypeptides

TITLE OF INVENTION: and Nucleic Acids

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/117,121

FILING DATE: 20-NOV-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/CA97/00062

FILING DATE: 30-JAN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 016252-00161005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:

LENGTH: 37 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: -

LOCATION: 1..37

OTHER INFORMATION: /note="HPLC-6"

US-09-117-121-39

Query Match 4.1%; Score 7; DB 4; Length 37;

Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 15

US-09-117-121-42

; Sequence 42, Application US/09117121

Patent No. 6307020

GENERAL INFORMATION:

APPLICANT: Hew, Choy

APPLICANT: Gong, Zhiyuan

TITLE OF INVENTION: Intracellular Antifreeze Polypeptides

TITLE OF INVENTION: and Nucleic Acids
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/117,121
FILING DATE: 20-NOV-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/CA97/00062
FILING DATE: 30-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 016252-00161005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-117-121-42

Query Match 4.1%; Score 7; DB 4; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
DB 6 AAAAAL 12

RESULT 16
US-09-344-529-4
Sequence 4, Application US/09344529
Patent No. 6429293
GENERAL INFORMATION:
APPLICANT: Hew, Choy L.
APPLICANT: HSC Research and Development Limited Partnership
TITLE OF INVENTION: Sculpin-Type Antifreeze Polypeptides and Nucleic Acids
FILE REFERENCE: 016252-00262005
CURRENT APPLICATION NUMBER: US/09/344,529
CURRENT FILING DATE: 1999-06-24
EARLIER APPLICATION NUMBER: US 60/090,794
EARLIER FILING DATE: 1998-06-26
EARLIER APPLICATION NUMBER: US 60/095,713
EARLIER FILING DATE: 1998-08-07
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 37
TYPE: PRT
ORGANISM: Pleuronectes americanus
FEATURE:
OTHER INFORMATION: Winter flounder liver-type antifreeze polypeptide
OTHER INFORMATION: (wLIAPF-6)
US-09-344-529-4

Query Match 4.1%; Score 7; DB 4; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
DB 6 AAAAAL 12

RESULT 17
US-07-814-220-1
Sequence 1, Application US/07814220
Patent No. 5925540
GENERAL INFORMATION:
APPLICANT: Caceci, Thomas
APPLICANT: Toth, Thomas E.
APPLICANT: Szumanski, Maria B.W.
TITLE OF INVENTION: SYNTHETIC ANTIFREEZE PEPTIDE AND
TITLE OF INVENTION: SYNTHETIC GENE CODING FOR ITS PRODUCTION
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: WHITHAM, CURTIS & WHITHAM
STREET: Reston Intl. Center, 11800 Sunrise Valley Dr.,
CITY: Reston
STATE: VA
COUNTRY: USA
ZIP: 20191
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/814,220
FILING DATE: 23-DEC-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/588,437
FILING DATE: 25-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Whitlam, Michael E.
REGISTRATION NUMBER: 32,635
REFERENCE/DOCKET NUMBER: CIT.016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-391-2510
TELEFAX: 703-391-9035
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-814-220-1

Query Match 4.1%; Score 7; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
DB 6 AAAAAL 12

RESULT 18
US-07-812-421-1
Sequence 1, Application US/07812421
Patent No. 5932697
GENERAL INFORMATION:
APPLICANT: Caceci, Thomas
APPLICANT: Toth, Thomas E.
APPLICANT: Szumanski, Maria B.W.
TITLE OF INVENTION: SYNTHETIC ANTIFREEZE PEPTIDE AND
TITLE OF INVENTION: SYNTHETIC GENE CODING FOR ITS PRODUCTION
US-07-812-421-1

```

; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WHITAM, CURTIS & WHITAM
; STREET: Reston Intl. Center, 11800 Sunrise Valley Dr.,
; CITY: Reston
; STATE: VA
; COUNTRY: USA
; ZIP: 20191
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/812,421
; FILING DATE: 23-DEC-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/588,437
; FILING DATE: 25-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlam, Michael E.
; REGISTRATION NUMBER: 32,635
; REFERENCE/DOCKET NUMBER: CTR.016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-391-2510
; TELEFAX: 703-391-9035
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-812-421-1

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAAL 14
Db 6 AAAAAAL 12

RESULT 19
US-09-325-932A-144
; Sequence 144, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Filinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: death and their use in the modification of forestry plant develo
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 144
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-325-932A-144

Query Match
Best Local Similarity 4.1%; Score 7; DB 4; Length 109;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAAL 14
Db 18 AAAAAAL 24
```

```

RESULT 20
US-08-233-609-3
; Sequence 3, Application US/08233609
; Patent No. 5534615
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,609
; FILING DATE: 25-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 894
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-233-609-3

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 203;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALSYL 17
Db 91 AALSYL 97

RESULT 21
US-08-444-083-3
; Sequence 3, Application US/08444083
; Patent No. 5571675
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
```

STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,083
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 894P1D5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-444-083-3

Query Match 4.1%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSVL 17
Db 91 AALSVL 97

RESULT 22
US-08-286-304-3
Sequence 3, Application US/08286304
Patent No. 5571893
GENERAL INFORMATION:
APPLICANT: Baker, Joffre
APPLICANT: Chien, Kenneth
APPLICANT: King, Kathleen
APPLICANT: Pennica, Diane
APPLICANT: Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,304
FILING DATE: 05-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994

ATTORNEY/AGENT INFORMATION:
NAME: Haasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 894P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-286-304-3

Query Match 4.1%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSVL 17
Db 91 AALSVL 97

RESULT 23
US-08-442-745-3
Sequence 3, Application US/08442745
Patent No. 5624806
GENERAL INFORMATION:
APPLICANT: Baker, Joffre
APPLICANT: Chien, Kenneth
APPLICANT: King, Kathleen
APPLICANT: Pennica, Diane
APPLICANT: Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/442,745
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 894P1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-442-745-3

Query Match 4.1%; Score 7; DB 1;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AALSVL 17
Db 91 AALSVL 97

RESULT 24

US-08-443-129-3
; Sequence 3, Application US/08443129
; Patent No. 5627073
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,129
; FILING DATE: 17-May-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/233609
; FILING DATE: 25-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/286
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: 304
; FILING DATE: (null)
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 894PID2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-443-129-3

Query Match 4.1%; Score 7; DB 1;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AALSVL 17
Db 91 AALSVL 97

RESULT 25
US-08-443-952-3
; Sequence 3, Application US/08443952

Patent No. 5679545
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,952
; FILING DATE: 17-May-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/233609
; FILING DATE: 25-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/286304
; FILING DATE: 05-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 894PID4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-443-952-3

Query Match 4.1%; Score 7; DB 1;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AALSVL 17
Db 91 AALSVL 97

RESULT 26
US-08-443-130-3
; Sequence 3, Application US/08443130
; Patent No. 5723585
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California

COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,130
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 894P1D3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-443-130-3

Query Match 4.1%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSVL 17
| | | | |
Db 91 AALSVL 97

RESULT 27
US-08-898-911-3
Sequence 3, Application US/08898911
Patent No. 6117650
GENERAL INFORMATION:
APPLICANT: King, Kathleen
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/898,911
FILING DATE: 23-Jul-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/452555

FILING DATE: 25-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P0894P1D6C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-08-898-911-3

Query Match 4.1%; Score 7; DB 3; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSVL 17
| | | | |
Db 91 AALSVL 97

RESULT 28
PCT-US95-04467-3
Sequence 3, Application PC/TUS9504467
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
TITLE OF INVENTION: Cardiotrophin and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04467
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 894P1PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US95-04467-3

Query Match 4.1%; Score 7; DB 5; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSVL 17
| | | | |
Db 91 AALSVL 97

RESULT 29
US-09-286-690-12
; Sequence 12, Application US/09286690
; Patent No. 6103511
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Lichense and Coding Sequences
; FILE REFERENCE: 55-96
; CURRENT APPLICATION NUMBER: US/09/286,690
; EARLIER FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,882
; EARLIER FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: PCT/US97/17811
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Fibrobacter succinogenes
US-09-286-690-12

Query Match
Best Local Similarity 4.1%; Score 7; DB 3; Length 228;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 15 AAAAAA 21

RESULT 30
US-08-508-761B-32
; Sequence 32, Application US/08508761B
; Patent No. 6027920
; GENERAL INFORMATION:
; APPLICANT: Joliff, Gwennael
; APPLICANT: Guyonvarch, Armel
; APPLICANT: Purification, Relano
; APPLICANT: Duchiron, Francis
; APPLICANT: Renaud, Michel
; TITLE OF INVENTION: System for Protein Expression and
; TITLE OF INVENTION: Secretion Especially in Corynebacteria
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
; STREET: 400 Seventh St. N.W.
; CITY: Washington D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,761B
; FILING DATE: 31-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/09652
; FILING DATE: 29-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/09870
; FILING DATE: 02-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Player, William E.
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P58525NA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666

TELEFAX: (202) 393-5350
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: M. kansasii
US-08-508-761B-32

Query Match
Best Local Similarity 4.1%; Score 7; DB 3; Length 309;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 3 AAAAAA 9

RESULT 31
US-08-107-676-30
; Sequence 30, Application US/08107676
; Patent No. 595356
; GENERAL INFORMATION:
; APPLICANT: Content, Jean
; APPLICANT: Demit, Lucas
; APPLICANT: Debryun, Jacqueline
; TITLE OF INVENTION: Mycobacterium Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding Them for Diagnosis and Control of
; TITLE OF INVENTION: Tuberculosis
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 595356west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,676
; FILING DATE: 04-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP92/00268
; FILING DATE: 07-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91400388.4
; FILING DATE: 14-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 8076.89USWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium kansasii
; IMMEDIATE SOURCE:
; CLONE: Partial protein sequence from antigen 85B
; CLONE: from M.kansasii

US-08-107-676-30

Query Match 4.1%; Score 7; DB 2; Length 325;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
111111
DB 19 AAAAAA 25

RESULT 32

US-09-161-241-9
; Sequence 9, Application US/09161241
; Patent No. 6344541
; GENERAL INFORMATION:
; APPLICANT: Bass, Michael B
; APPLICANT: Sullivan, John K
; APPLICANT: Theill, Lars E
; APPLICANT: Wang, Daquan
; TITLE OF INVENTION: NOVEL DKR POLYPEPTIDES
; FILE REFERENCE: A-548
; CURRENT APPLICATION NUMBER: US/09/161,241
; CURRENT FILING DATE: 1998-09-25
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Human
US-09-161-241-9

Query Match 4.1%; Score 7; DB 4; Length 350;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
111111
DB 338 AAAAAA 344

RESULT 33

US-07-792-885A-1
; Sequence 1, Application US/07792885A
; Patent No. 551651
; GENERAL INFORMATION:
; APPLICANT: Goldring, Steven R.
; APPLICANT: Gorn, Alan H.
; APPLICANT: Lin, Herb Y.
; TITLE OF INVENTION: MAMMALIAN CALCITONIN RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,885A
; FILING DATE: 19911115
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00786/074001

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

LENGTH: 482
; TYPE: AMINO ACID
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
US-07-792-885A-1

Query Match 4.1%; Score 7; DB 1; Length 482;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
111111
DB 431 AAAAAA 437

RESULT 34

US-08-142-439A-7
; Sequence 7, Application US/08142439A
; Patent No. 5670360
; GENERAL INFORMATION:
; APPLICANT: Thorens, Bernard
; TITLE OF INVENTION: Receptor for the Glucagon-Like-Peptide-1
; TITLE OF INVENTION: (GLP-1)
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56703600 No. 56703600 of No. 56703600th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/142,439A

FILING DATE: 24-NOV-93

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 398/92

FILING DATE: 25-MAR-92

PRIOR APPLICATION DATA: PCT/EP93/00697

APPLICATION NUMBER: 23-MAR-93

FILING DATE: 23-MAR-93

ATTORNEY/AGENT INFORMATION:

NAME: Harrington, James J.

REGISTRATION NUMBER: 38,711

REFERENCE/DOCKET NUMBER: 3756, 204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 482 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Sus scrofa

US-08-142-439A-7

Query Match 4.18; Score 7; DB 1; Length 482;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
Db 431 AAAAAL 437

RESULT 35
US-08-869-477-7
Sequence 7; Application US/08869477
Patent No. 5846747
GENERAL INFORMATION:
APPLICANT: Thoenes, Bernard
TITLE OF INVENTION: Receptor for the Glucagon-Like-Peptide-1
TITLE OF INVENTION: (GLP-1)
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5846747/0 No. 5846747/disk of No. 5846747th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,477
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/142,439
FILING DATE: 24-NOV-93
APPLICATION NUMBER: DK 398/92
FILING DATE: 25-MAR-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/00697
FILING DATE: 23-MAR-93
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 3756.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 482 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
US-08-869-477-7

Query Match 4.18; Score 7; DB 2; Length 482;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
Db 431 AAAAAL 437

RESULT 36
US-09-172-952-26

Sequence 26; Application US/09172952
Patent No. 6368793
GENERAL INFORMATION:
APPLICANT: Hoch, James
APPLICANT: Dattois, Veronique
TITLE OF INVENTION: METABOLIC SELECTION METHODS
FILE REFERENCE: 234/191
CURRENT APPLICATION NUMBER: US/09/172,952
CURRENT FILING DATE: 1998-10-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 26
LENGTH: 498
TYPE: PRT
ORGANISM: Lyxk-Ec
US-09-172-952-26

Query Match 4.18; Score 7; DB 4; Length 498;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 YNMSLGE 122
Db 184 YNMSLGE 190

RESULT 37
US-09-382-256-10
Sequence 10; Application US/09382256A
Patent No. 6207814
GENERAL INFORMATION:
APPLICANT: MIYAZONO, Kohel
TEN DIKE, Peter
FRANZEN, Petra
YAMASHITA, Hidetoshi
HELDIN, Carl-Henrik
TITLE OF INVENTION: ACTIVIN RECEPTOR LIKE KINASES, PROTEINS
HAVING SERINE THREONINE KINASE DOMAINS,
AND THEIR USE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/382,256A
FILING DATE: 24-Aug-1999
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: No. 6207814ember 17, 1993
APPLICATION NUMBER: GB 9224057.1
FILING DATE: No. 6207814ember 17, 1992
APPLICATION NUMBER: GB 9304677.9
FILING DATE: March 8, 1993
APPLICATION NUMBER: GB 9304680.3
FILING DATE: March 8, 1993
APPLICATION NUMBER: 9311047.6
FILING DATE: May 28, 1993
APPLICATION NUMBER: 9313763.6
FILING DATE: July 2, 1993
APPLICATION NUMBER: 9316099.2
FILING DATE: August 3, 1993
APPLICATION NUMBER: 321344.5
FILING DATE: October 15, 1993

ATTORNEY/AGENT INFORMATION:
NAME: NO. 6207814man D. Hanson
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5298.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-382-256-10

Query Match 4.1%; Score 7; DB 4; Length 503;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
| | | | |
Db 21 AAAAAA 27

RESULT 38
US-09-395-115-10
Sequence 10, Application US/09395115
Patent No. 6271365

GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; Dijke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/395,115
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6271365ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6271365ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohel, Vlnet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-395-115-10

Query Match 4.1%; Score 7; DB 4; Length 503;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
| | | | |
Db 21 AAAAAA 27

RESULT 39
US-08-436-265-10
Sequence 10, Application US/08436265
Patent No. 6316217

GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; Dijke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6316217ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6316217ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
TELECOMMUNICATION INFORMATION: LUD 5298
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-265-10

Query Match 4.1%; Score 7; DB 4; Length 503;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
DB 21 AAAAAL 27

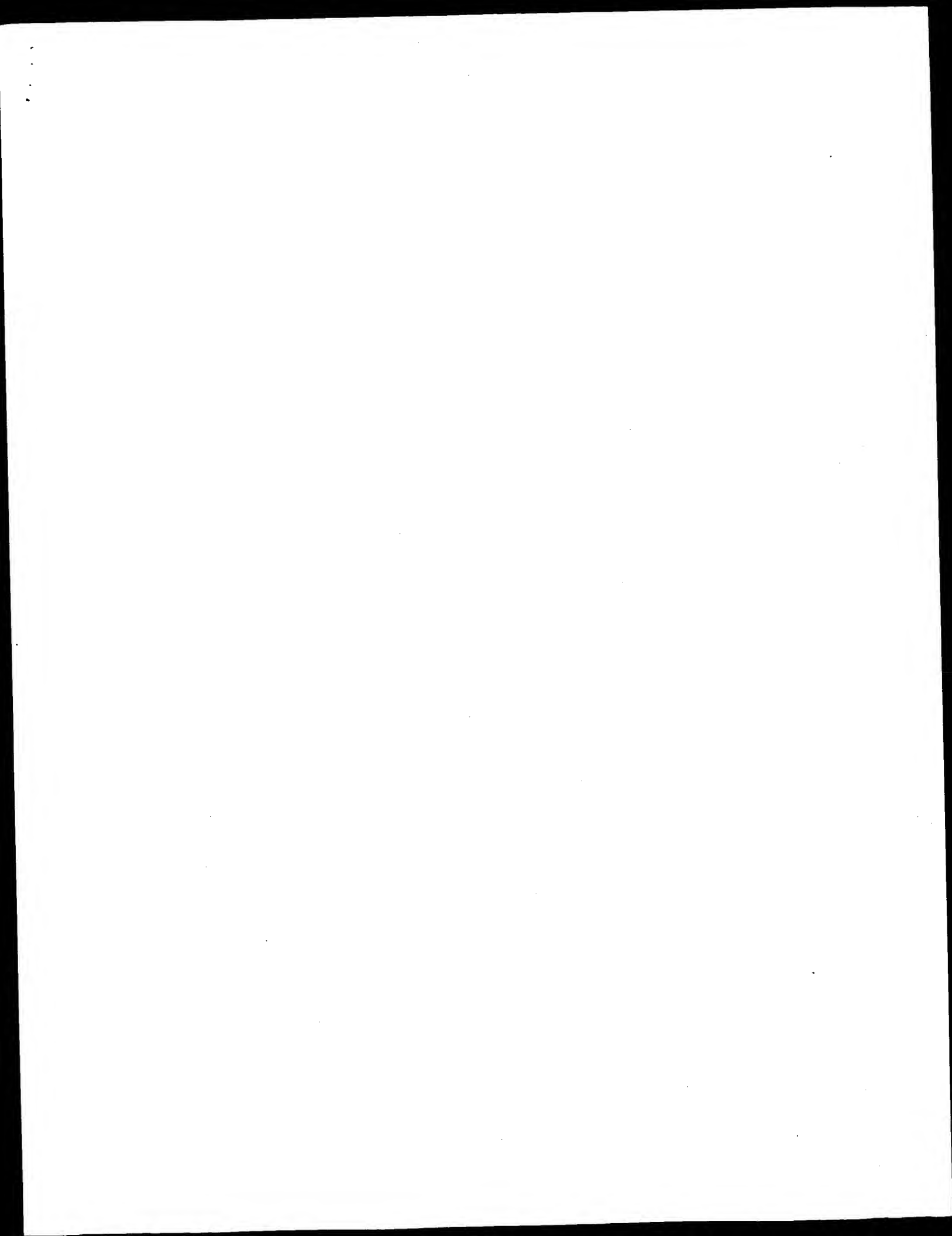
RESULT 40
US-09-679-187-10
Sequence 10, Application US/09679187
Patent No. 6331621
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohlei, Dijke, Peter Ten;
TITLE OF INVENTION: Fransen, Petra; Yamashita, Hidetoshi; Hejdin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/679,187
FILING DATE: 03-October-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-NO. 6331621ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-NO. 6331621ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
TELECOMMUNICATION INFORMATION: LUD 5298
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-679-187-10

Query Match 4.1%; Score 7; DB 4; Length 503;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
DB 21 AAAAAL 27

Search completed: July 6, 2003, 14:22:34
Job time: 21 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:22:40 ; Search time 24 Seconds

(without alignments)
824.116 Million cell updates/sec

Title: US-09-674-779b-2

Perfect score: 172
Sequence: 1 MMLHIQIAAAALSVLTFM.....IAFGTNEAMSONRAELSY 172

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 440863 seqs, 114992915 residues

Word size : 0

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PTI_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	4.7	1744	9 US-10-108-605-25	Sequence 25, Appl
2	7	4.1	21	10 US-09-864-761-42679	Sequence 42679, A
3	7	4.1	22	9 US-10-024-450-7	Sequence 7, Appl
4	7	4.1	23	9 US-10-311-111-29	Sequence 29, Appl
5	7	4.1	38	9 US-09-776-724A-165	Sequence 165, App
6	7	4.1	41	9 US-10-106-487-1	Sequence 1, Appl
7	7	4.1	52	10 US-09-864-761-40439	Sequence 40439, A
8	7	4.1	78	10 US-09-795-501-10	Sequence 10, Appl
9	7	4.1	79	9 US-10-092-154-770	Sequence 770, App
10	7	4.1	109	9 US-09-764-847-770	Sequence 770, App
11	7	4.1	119	9 US-10-219-220-144	Sequence 144, App
12	7	4.1	131	10 US-09-764-891-2737	Sequence 2737, Ap
13	7	4.1	132	10 US-09-164-615-44	Sequence 44, Appl
14	7	4.1	132	10 US-09-164-615-49	Sequence 49, Appl
15	7	4.1	161	10 US-09-864-761-42364	Sequence 42364, A
16	7	4.1	182	10 US-09-747-348-2	Sequence 2, Appl
17	7	4.1	203	9 US-10-107-931-3	Sequence 3, Appl
18	7	4.1	203	10 US-09-896-856-3	Sequence 3, Appl
19	7	4.1	273	10 US-09-864-761-36691	Sequence 36691, A

20	7	4.1	287	9 US-10-156-761-9684	Sequence 9684, Ap
21	7	4.1	300	9 US-10-156-761-12279	Sequence 12279, A
22	7	4.1	319	9 US-10-156-761-9015	Sequence 9015, Ap
23	7	4.1	324	9 US-10-281-024-15	Sequence 15, Appl
24	7	4.1	340	9 US-10-156-761-9858	Sequence 9858, Ap
25	7	4.1	350	9 US-09-905-291A-236	Sequence 236, Appl
26	7	4.1	350	9 US-09-976-736-9	Sequence 9, Appl
27	7	4.1	350	9 US-10-063-547-8	Sequence 8, Appl
28	7	4.1	350	9 US-09-902-853-236	Sequence 236, App
29	7	4.1	350	9 US-09-907-824-236	Sequence 236, App
30	7	4.1	350	9 US-09-907-841-236	Sequence 236, App
31	7	4.1	350	9 US-09-904-011-236	Sequence 236, App
32	7	4.1	350	9 US-10-201-310-2	Sequence 2, Appl
33	7	4.1	350	9 US-10-063-616-8	Sequence 8, Appl
34	7	4.1	350	9 US-09-906-742-236	Sequence 236, App
35	7	4.1	350	9 US-10-063-502-8	Sequence 8, Appl
36	7	4.1	350	9 US-09-906-838-236	Sequence 236, App
37	7	4.1	350	9 US-09-907-613-236	Sequence 236, App
38	7	4.1	350	9 US-09-907-942-236	Sequence 236, App
39	7	4.1	350	9 US-09-904-820-236	Sequence 236, App
40	7	4.1	350	9 US-09-904-859-236	Sequence 236, App
41	7	4.1	350	9 US-09-909-204-236	Sequence 236, App
42	7	4.1	350	9 US-09-904-786-236	Sequence 236, App
43	7	4.1	350	9 US-09-906-646-236	Sequence 236, App
44	7	4.1	350	9 US-09-906-700-236	Sequence 236, App
45	7	4.1	350	9 US-09-902-903-236	Sequence 236, App
46	7	4.1	350	9 US-09-903-749A-236	Sequence 236, App
47	7	4.1	350	9 US-09-903-786-236	Sequence 236, App
48	7	4.1	350	9 US-09-902-736-236	Sequence 236, App
49	7	4.1	350	9 US-09-904-119-236	Sequence 236, App
50	7	4.1	350	9 US-09-904-956-236	Sequence 236, App
51	7	4.1	350	9 US-09-907-794-236	Sequence 236, App
52	7	4.1	350	9 US-10-063-518-8	Sequence 8, Appl
53	7	4.1	350	9 US-10-063-598-8	Sequence 8, Appl
54	7	4.1	350	9 US-10-227-693-8	Sequence 8, Appl
55	7	4.1	350	9 US-09-902-692-236	Sequence 236, App
56	7	4.1	350	9 US-09-903-520-236	Sequence 236, App
57	7	4.1	350	9 US-09-903-943-236	Sequence 236, App
58	7	4.1	350	9 US-09-904-462-236	Sequence 236, App
59	7	4.1	350	9 US-09-905-056-236	Sequence 236, App
60	7	4.1	350	9 US-09-907-925-236	Sequence 236, App
61	7	4.1	350	9 US-09-904-553-236	Sequence 236, App
62	7	4.1	350	9 US-09-905-381-236	Sequence 236, App
63	7	4.1	350	9 US-09-909-064-236	Sequence 236, App
64	7	4.1	350	9 US-09-972-473-2	Sequence 2, Appl
65	7	4.1	350	9 US-10-063-567-8	Sequence 8, Appl
66	7	4.1	350	9 US-09-905-088-236	Sequence 236, App
67	7	4.1	350	9 US-09-907-575-236	Sequence 236, App
68	7	4.1	350	9 US-10-063-538-8	Sequence 8, Appl
69	7	4.1	350	9 US-09-902-759-236	Sequence 236, App
70	7	4.1	350	9 US-09-905-075-236	Sequence 236, App
71	7	4.1	350	9 US-10-063-599-8	Sequence 8, Appl
72	7	4.1	350	9 US-09-902-634-236	Sequence 236, App
73	7	4.1	350	9 US-09-902-713-236	Sequence 236, App
74	7	4.1	350	9 US-09-907-979-236	Sequence 236, App
75	7	4.1	350	9 US-10-063-595-8	Sequence 8, Appl
76	7	4.1	350	9 US-09-902-615-236	Sequence 236, App
77	7	4.1	350	9 US-09-903-925-236	Sequence 236, App
78	7	4.1	350	9 US-09-906-760A-236	Sequence 236, App
79	7	4.1	350	9 US-10-223-088-50	Sequence 50, Appl
80	7	4.1	350	9 US-10-271-628-3	Sequence 3, Appl
81	7	4.1	350	9 US-09-907-652-236	Sequence 236, App
82	7	4.1	350	9 US-09-903-823-236	Sequence 236, App
83	7	4.1	350	9 US-10-063-580-8	Sequence 8, Appl
84	7	4.1	350	9 US-10-223-084-50	Sequence 50, Appl
85	7	4.1	350	9 US-10-223-088-50	Sequence 50, Appl
86	7	4.1	350	9 US-10-223-090-50	Sequence 50, Appl
87	7	4.1	350	9 US-09-902-572A-236	Sequence 236, App
88	7	4.1	350	9 US-10-063-557-8	Sequence 8, Appl
89	7	4.1	350	9 US-10-223-087-50	Sequence 50, Appl
90	7	4.1	350	9 US-09-902-979-236	Sequence 236, App
91	7	4.1	350	9 US-09-905-125-236	Sequence 236, App
92	7	4.1	350	9 US-09-906-815A-236	Sequence 236, App

93 7 4.1 350 9 US-10-063-585-8 Sequence 8, Appl
 94 7 4.1 350 9 US-10-223-083-50 Sequence 50, Appl
 95 7 4.1 350 10 US-09-909-320-236 Sequence 236, App
 96 7 4.1 350 10 US-09-909-088B-236 Sequence 236, App
 97 7 4.1 350 12 US-10-006-867-8 Sequence 8, Appl
 98 7 4.1 356 10 US-09-815-242-4875 Sequence 4875, Ap
 99 7 4.1 356 10 US-09-815-242-10782 Sequence 10782, A
 100 7 4.1 395 9 US-10-153-668-330 Sequence 330, App

ALIGNMENTS

RESULT 1
 US-10-108-605-25
 ; Sequence 25, Application US/10108605
 ; Patent No. US20020160934A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Broadus, Julie
 ; APPLICANT: Stamm, Lynn Jane
 ; APPLICANT: Bachmann, Kim
 ; APPLICANT: Kanda, Kim
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
 ; FILE REFERENCE: 31133B
 ; PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
 ; CURRENT APPLICATION NUMBER: US/10/108, 605
 ; PRIOR FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: US 09/761,142
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/176,418
 ; PRIOR FILING DATE: 2000-01-14
 ; NUMBER OF SEQ ID NOS: 361
 ; SOFTWARE: Patent Ver. 2.1
 ; SEQ ID NO 25
 ; LENGTH: 1744
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; US-10-108-605-25

Query Match 4.7%; Score 8; DB 9; Length 1744;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAALS 15
 |||||
 Db 941 AAAAAALS 948

RESULT 2
 US-09-864-761-42679
 ; Sequence 42679, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecmca-x-1
 ; GENE EXPRESSION ANALYSIS BY MICROARRAY
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; PRIOR FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263, 6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 42679
 ; LENGTH: 21
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC007567.1
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.7
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.86
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.3
 ; OTHER INFORMATION: EST_HUMAN HIT: AL134791.1, EVALUATE 1.00e-03
 ; US-09-864-761-42679

Query Match 4.1%; Score 7; DB 10; Length 21;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 NYLKGK 136
 |||||
 Db 4 NYLKGK 10

RESULT 3
 US-10-024-450-7
 ; Sequence 7, Application US/10024450
 ; Publication No. US20030032606A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Huang, Shi
 ; APPLICANT: Chadwick, Robert B.
 ; TITLE OF INVENTION: Methods of Detecting and Treating
 ; FILE REFERENCE: P-11 5101
 ; Microsatellite-Instability Positive Tumors Using RIZ
 ; CURRENT APPLICATION NUMBER: US/10/024,450
 ; PRIOR FILING DATE: 2001-12-17
 ; PRIOR APPLICATION NUMBER: US 60/256,582
 ; PRIOR FILING DATE: 2000-12-19
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 22
 ; TYPE: PRT
 ; ORGANISM: Homo sapien

US-10-024-450-7

Query Match

Best Local Similarity 4.1%; Score 7; DB 9; Length 22;
Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 4

US-10-311-111-29

Sequence 29, Application US/10311111
Publication No. US20030121065A1

GENERAL INFORMATION:

APPLICANT: SHIBA, KIYOTAKA

TITLE OF INVENTION: MULTIFUNCTIONAL BASE SEQUENCE AND ARTIFICIAL GENE CONTAINING THE

FILE REFERENCE: 4439-4004

CURRENT FILING DATE: 2002-12-13

PRIOR APPLICATION NUMBER: JP 2000-180997

PRIOR FILING DATE: 2000-06-16

NUMBER OF SEQ ID NOS: 34

SOFTWARE: PatentIn version 3.1

SEQ ID NO 29

LENGTH: 23

TYPE: PRT

ORGANISM: artificial

FEATURE:

OTHER INFORMATION: Designed peptide

US-10-311-111-29

Query Match

Best Local Similarity 4.1%; Score 7; DB 9; Length 23;
Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 5

US-09-776-724A-165

Sequence 165, Application US/09776724A

Publication No. US20030050455A1

GENERAL INFORMATION:

APPLICANT: ROSEN, et al.

TITLE OF INVENTION: 64 Human Secreted Proteins

FILE REFERENCE: P2011

CURRENT FILING DATE: 2001-02-06

PRIOR APPLICATION NUMBER: 60/180,909

PRIOR FILING DATE: 2000-02-08

PRIOR APPLICATION NUMBER: 09/669,688

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: 09/229,982

PRIOR FILING DATE: 1999-01-14

PRIOR APPLICATION NUMBER: PCT/US98/14613

PRIOR FILING DATE: 1998-07-15

PRIOR APPLICATION NUMBER: 60/052,661

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,872

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,871

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,874

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,873

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,870

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,875

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/053,440

PRIOR FILING DATE: 1997-07-22

PRIOR APPLICATION NUMBER: 60/053,441

PRIOR FILING DATE: 1997-07-22

PRIOR APPLICATION NUMBER: 60/053,442

PRIOR FILING DATE: 1997-07-22

PRIOR APPLICATION NUMBER: 60/056,359

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,725

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,985

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,952

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,989

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/056,361

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,726

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,724

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,946

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,683

PRIOR FILING DATE: 1997-08-18

NUMBER OF SEQ ID NOS: 294

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 165

LENGTH: 38

TYPE: PRT

ORGANISM: Homo sapiens

US-09-776-724A-165

Query Match

Best Local Similarity 4.1%; Score 7; DB 9; Length 38;
Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 8 AAAAAA 14
Db 23 AAAAAA 29RESULT 6
US-10-106-487-1
Sequence 1, Application US/10106487

Patent No. US20020164721A1

GENERAL INFORMATION:

APPLICANT: FIRAT, HOSEYIN

APPLICANT: LEMONNIER, FRANCOIS

APPLICANT: LANGIARDE-DEMOYEN, PIERRE

APPLICANT: MICHEL, MARIE-LOUISE

TITLE OF INVENTION: DESIGN OF A POLYPEPTIDIC CONSTRUCT FOR THE INDUCTION

TITLE OF INVENTION: OF HLA-A2.1 RESTRICTED HIV 1 SPECIFIC CTL RESPONSES USING

TITLE OF INVENTION: HHD MICE

FILE REFERENCE: 03495.0196 SEQUENCE LISTING

CURRENT APPLICATION NUMBER: US/10/106,487

CURRENT FILING DATE: 2002-03-27

PRIOR APPLICATION NUMBER: 09/675,673

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/158,356

PRIOR FILING DATE: 1999-10-12

NUMBER OF SEQ ID NOS: 41

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 41

TYPE: PRT

ORGANISM: Hepatitis B virus

US-10-106-487-1

Query Match

4.1%; Score 7; DB 9; Length 41;

Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 7

US-09-864-761-40439
; Sequence 40439, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40439
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121747.19
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: EST_HUMAN HIT: A0131227.1, EVALUATE 3.00e-05
US-09-864-761-40439

Query Match 4.1%; Score 7; DB 10; Length 52;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
Db 7 AAAAAA 13

RESULT 8

US-09-795-501-10
; Sequence 10, Application US/09795501
; Patent No. US20020042098A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32253 TRANSFERASE FAMILY MEMBERS AND
; FILE REFERENCE: 38155-20004.00
; CURRENT FILING DATE: 2001-07-28
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/185,755
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid sequence
US-09-795-501-10

Query Match 4.1%; Score 7; DB 10; Length 78;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
Db 60 AAAAAA 66

RESULT 9

US-10-092-154-770
; Sequence 770, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT FILING DATE: 2002-03-07
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See file wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 770
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-092-154-770

Query Match 4.1%; Score 7; DB 9; Length 79;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAA 13
Db 42 IAAAAA 48

RESULT 10
US-09-764-847-770
; Sequence 770, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 770
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-847-770

Query Match
Best Local Similarity 4.1%; Score 7; DB 10; Length 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAA 13
Db 42 IAAAAA 48

RESULT 11
US-10-219-220-144
; Sequence 144, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 144
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-10-219-220-144

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 109;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 18 AAAAAA 24

RESULT 12
US-09-764-891-2737
; Sequence 2737, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2737
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (103)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-2737

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 119;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 77 AAAAAA 83

RESULT 13
US-09-164-615-44
; Sequence 44, Application US/09164615B
; Patent No. US20020073442A1
; GENERAL INFORMATION:
; APPLICANT: Hanley-Bowdoin, Linda
; APPLICANT: Settlage, Sharon
; TITLE OF INVENTION: Geminivirus Resistant Transgenic Plants
; FILE REFERENCE: 5051-433
; CURRENT APPLICATION NUMBER: US/09/164,615B
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TYCLV mutant
US-09-164-615-44

Query Match
Best Local Similarity 4.1%; Score 7; DB 10; Length 131;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAA 13
Db 48 IAAAAA 54

RESULT 14
US-09-164-615-49
; Sequence 49, Application US/09164615B
; Patent No. US20020073442A1
; GENERAL INFORMATION:
; APPLICANT: Hanley-Bowdoin, Linda
; APPLICANT: Settlage, Sharon
; TITLE OF INVENTION: Geminivirus Resistant Transgenic Plants
; FILE REFERENCE: 5051-433
; CURRENT APPLICATION NUMBER: US/09/164,615B
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 132

```
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: TGMV AL3
;; OTHER INFORMATION: mutant (MAL3#67)
US-09-164-615-49

Query Match
Best Local Similarity 4.1%; Score 7; DB 10; Length 132;
100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 49 AAAAAA 55

RESULT 15
US-09-864-761-42364
; Sequence 42364, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42364
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
;; OTHER INFORMATION: MAP TO AC011447.2
;; OTHER INFORMATION: EXPRESSED IN PLECENTRA, SIGNAL = 3.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.8
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.9
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: SWISSPROT HIT: P11227, EVALUATE 9.00e-14
;; OTHER INFORMATION: EST_HUMAN HIT: AM753028.1, EVALUATE 7.00e-61
US-09-864-761-42364

Query Match
Best Local Similarity 4.1%; Score 7; DB 10; Length 161;
100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 7 AAAAAA 13

RESULT 16
US-09-747-348-2
; Sequence 2, Application US/09747348
; Patent No. US20020123067A1
; GENERAL INFORMATION:
; APPLICANT: Andrew D. MURDIN, Raymond P. OOMEN, Joe WANG, Pamela DUNN
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses th
; FILE REFERENCE: 77813-37
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 60/171,525
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-747-348-2

Query Match
Best Local Similarity 4.1%; Score 7; DB 10; Length 192;
100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 GHTDERG 112
Db 123 GHTDERG 129

RESULT 17
US-10-107-931-3
; Sequence 3, Application US/10107931
; Publication No. US20030054550A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/107,931
FILING DATE: 26-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/364,187
FILING DATE: 30-Jul-1999
APPLICATION NUMBER: 08/286,304
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 894P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-107-931-3

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 203;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALSVL 17
Db 91 AAALSVL 97

RESULT 18
US-09-896-856-3
Sequence 3, Application US/09896856
Patent No. US20020137189A1
GENERAL INFORMATION:
APPLICANT: Baker, Joffe
Chien, Kenneth
King, Kathleen
Pennica, Diane
Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/896,856
FILING DATE: 29-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/733,850
FILING DATE: 18-OCT-1996
APPLICATION NUMBER: US 08/471,112
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: 08/233,609
FILING DATE: 25-APR-1994
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994

APPLICATION NUMBER: 08/443129
FILING DATE: 17-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: F0894PID2C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-2066
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-896-856-3

Query Match
Best Local Similarity 4.1%; Score 7; DB 10; Length 203;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAALSVL 17
Db 91 AAALSVL 97

RESULT 19
US-09-864-761-36691
Sequence 36691, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
Applicant: Rank, David R.
Applicant: Hanzel, David K.
Applicant: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Acomics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21

```

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36691
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006460.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BT4/4, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: SWISSPROT HIT: 062722, EVALU0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: A0131373.1, EVALU0.100e-114
; US-09-B64-761-36691

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 273;
Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 10 AAAALSV 16
    |||||
Db 193 AAAALSV 199

```

```

RESULT 20
US-10-156-761-9684
; Sequence 9684, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9684
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-9684

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 287;
Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 8 AAAAAL 14
    |||||
Db 157 AAAAAL 163

```

```

RESULT 21
US-10-156-761-12279

```

```

; Sequence 12279, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12279
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-12279

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 300;
Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 8 AAAAAL 14
    |||||
Db 43 AAAAAL 49

```

```

RESULT 22
US-10-156-761-9015
; Sequence 9015, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9015
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-9015

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 319;
Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 9 AAAAALS 15
    |||||
Db 301 AAAAALS 307

```

```

RESULT 23
US-10-281-024-15
; Sequence 15, Application US/10281024

```

```

; Publication No. US20030087410A1
; GENERAL INFORMATION:
; APPLICANT: SATOSHI MORI
; APPLICANT: KYOKO HIGUCHI
; TITLE OF INVENTION: NICOTINAMINE SYNTHASE AND GENE ENCODING
; FILE REFERENCE: THE SAME
; CURRENT APPLICATION NUMBER: US/10/281,024
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US/09/674,337A
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: PCT/JP99/02305
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Oryza sativa L.
US-10-281-024-15

```

```

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 324;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 63 ASKLPSTL 69
Db 20 ASKLPSTL 26

```

```

RESULT 24
US-10-156-761-9858
; Sequence 9858, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9858
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9858

```

```

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 340;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 122 ERRAVAV 128
Db 199 ERRAVAV 205

```

```

RESULT 25
US-09-905-291A-236
; Sequence 236, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi

```

```

; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Goddard, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gunney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Peoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 236
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-291A-236

```

```

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 8 AAAAAA 14
 |||||
 Db 338 AAAAAA 344

RESULT 26
 US-09-976-736-9
 ; Sequence 9, Application US/09976736
 ; Patent No. US20020161178A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bass, Michael B
 ; APPLICANT: Sullivan, John K
 ; APPLICANT: Theill, Lars E
 ; APPLICANT: Wang, Daquan
 ; TITLE OF INVENTION: NOVEL DKR POLYPEPTIDES
 ; FILE REFERENCE: A-548
 ; CURRENT APPLICATION NUMBER: US/09/976,736
 ; CURRENT FILING DATE: 2001-10-09
 ; PRIOR APPLICATION NUMBER: US/09/161,241
 ; PRIOR FILING DATE: 1998-09-25
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 9
 ; LENGTH: 350
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-976-736-9

Query Match 4.1%; Score 7; DB 9; Length 350;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 |||||
 Db 338 AAAAAA 344

RESULT 27
 US-10-063-547-8
 ; Sequence 8, Application US/10063547
 ; Publication No. US20020182638A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gutney, Austin L.
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3230R1C1
 ; CURRENT APPLICATION NUMBER: US/10/063,547
 ; CURRENT FILING DATE: 2002-05-02
 ; PRIOR APPLICATION removed - See file Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 170
 ; SEQ ID NO 8
 ; LENGTH: 350
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-063-547-8

Query Match 4.1%; Score 7; DB 9; Length 350;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 |||||
 Db 338 AAAAAA 344

RESULT 28
 US-09-902-853-236
 ; Sequence 236, Application US/09902853
 ; Publication No. US20020192659A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, A.
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gutney, Austin L.
 ; APPLICANT: Hillan, Kenneth, J.
 ; APPLICANT: Kijavlin, Ivar J.
 ; APPLICANT: Mather, Jennie P.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Williams, P. Mickey
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ; TITLE OF INVENTION: Acids Encoding the Same
 ; FILE REFERENCE: 10466-14
 ; CURRENT APPLICATION NUMBER: US/09/902,853
 ; CURRENT FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: US/09/665,350
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: US 60/143,048
 ; PRIOR FILING DATE: 1999-07-07
 ; PRIOR APPLICATION NUMBER: US 60/145,698
 ; PRIOR FILING DATE: 1999-07-26
 ; PRIOR APPLICATION NUMBER: US 60/146,222
 ; PRIOR FILING DATE: 1999-07-28
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-08
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-13
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313
 ; PRIOR FILING DATE: 1999-11-30
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 236
 ; LENGTH: 350
 ; TYPE: PRT

ORGANISM: Homo Sapien
US-09-902-853-236

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 338 AAAAAA 344

RESULT 29

US-09-907-824-236
Sequence 236, Application US/09907824
Publication No. US2002019671A1
GENERAL INFORMATION:

APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertsens, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gunney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,824
PRIOR FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Sapien
US-09-907-824-236

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 338 AAAAAA 344

RESULT 30

US-09-907-841-236
Sequence 236, Application US/09907841
Publication No. US20020198366A1
GENERAL INFORMATION:

APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertsens, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gunney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,841
PRIOR FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13

```
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; LENGTH: 350
; SEQ ID NO 236
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-841-236

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

OY 8 AAAAAA 14
DB 338 AAAAAA 344

RESULT 31
US-09-904-011-236
; Sequence 236, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kiljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
```

```
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 236
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-904-011-236

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

OY 8 AAAAAA 14
DB 338 AAAAAA 344

RESULT 32
US-10-201-310-2
; Sequence 2, Application US/10201310
; Publication No. US20030004327A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Ruden, Steven M.
; TITLE OF INVENTION: Cerebellum and Embryo Specific Protein
; FILE REFERENCE: 1488,0610002
; CURRENT APPLICATION NUMBER: US/10/201,310
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US 60/033,870
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 08/993,198
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-201-310-2

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

OY 8 AAAAAA 14
DB 338 AAAAAA 344

RESULT 33
US-10-063-616-8
```

```
Sequence 8, Application US/10063616
Publication No. US20030013855A1
GENERAL INFORMATION:
APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063,616
PRIOR FILING DATE: 2002-05-03
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 8
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Saplen
US-10-063-616-8

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
Db 338 AAAAAAL 344

RESULT 34
US-09-906-742-236
Sequence 236, Application US/09906742
Publication No. US20030023054A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kiljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paonl, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/906,742
PRIOR FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
ORGANISM: Homo Saplen
```

```
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Saplen
US-09-906-742-236

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
Db 338 AAAAAAL 344

RESULT 35
US-10-063-502-8
Sequence 8, Application US/10063502
Publication No. US20030023042A1
GENERAL INFORMATION:
APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063,502
CURRENT FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: 09/665,350
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 8
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Saplen
```

US-10-063-502-8

Query Match 4.1% Score 7; DB 9; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
|||||

DB 338 AAAAAL 344

RESULT 36

US-09-906-838-236
Sequence 236, Application US/09906838
Publication No. US20030027143A1

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gertlisen, Mary E.

APPLICANT: Goddard, A.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillen, Kenneth, J.

APPLICANT: Kijavlin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tamas, Daniel

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

FILE REFERENCE: 10466-14

CURRENT APPLICATION NUMBER: US/09/906,838

CURRENT FILING DATE: 2001-07-16

PRIOR APPLICATION NUMBER: 09/665,350

PRIOR FILING DATE: 2000-09-18

PRIOR APPLICATION NUMBER: PCT/US00/04414

PRIOR FILING DATE: 2000-02-22

PRIOR APPLICATION NUMBER: US 60/143,048

PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: US 60/145,698

PRIOR FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: US 60/146,222

PRIOR FILING DATE: 1999-07-28

PRIOR APPLICATION NUMBER: PCT/US99/20594

PRIOR FILING DATE: 1999-09-08

PRIOR APPLICATION NUMBER: PCT/US99/20944

PRIOR FILING DATE: 1999-09-13

PRIOR APPLICATION NUMBER: PCT/US99/21090

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/21547

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/23089

PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: PCT/US99/28214

PRIOR FILING DATE: 1999-11-29

PRIOR APPLICATION NUMBER: PCT/US99/28313

PRIOR FILING DATE: 1999-11-30

PRIOR APPLICATION NUMBER: PCT/US99/28564

PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: PCT/US99/28565

PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: PCT/US99/30095

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: PCT/US99/30911

PRIOR FILING DATE: 1999-12-20

PRIOR APPLICATION NUMBER: PCT/US99/30999

PRIOR FILING DATE: 1999-12-20

PRIOR APPLICATION NUMBER: PCT/US00/00219

PRIOR FILING DATE: 2000-01-05

NUMBER OF SEQ ID NOS: 423

SEQ ID NO 236

LENGTH: 350

TYPE: PRT

ORGANISM: Homo Sapien

US-09-906-838-236

Query Match 4.1% Score 7; DB 9; Length 350;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
|||||

DB 338 AAAAAL 344

RESULT 37

US-09-907-613-236
Sequence 236, Application US/09907613
Publication No. US20030027145A1

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gertlisen, Mary E.

APPLICANT: Goddard, A.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillen, Kenneth, J.

APPLICANT: Kijavlin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tamas, Daniel

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

FILE REFERENCE: 10466-14

CURRENT APPLICATION NUMBER: US/09/907,613

CURRENT FILING DATE: 2001-07-17

PRIOR APPLICATION NUMBER: PCT/US00/04414

PRIOR FILING DATE: 2000-02-22

PRIOR APPLICATION NUMBER: US 60/143,048

PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: US 60/145,698

PRIOR FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: PCT/US99/20594

PRIOR FILING DATE: 1999-07-28

PRIOR APPLICATION NUMBER: PCT/US99/20944

PRIOR FILING DATE: 1999-09-08

PRIOR APPLICATION NUMBER: PCT/US99/20944

PRIOR FILING DATE: 1999-09-13

PRIOR APPLICATION NUMBER: PCT/US99/21090

```

; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 236
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-613-236

```

```

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 8 AAAAAA 14
Db 338 AAAAAA 344

```

RESULT 38

```

; Sequence 236, Application US/09907942
; Publication No. US20030027146A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Bolstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907/942
; PRIOR APPLICATION NUMBER: PCT/US00/04414

```

Query Match

```

Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 8 AAAAAA 14
Db 338 AAAAAA 344

```

RESULT 39

```

; Sequence 236, Application US/09904820
; Publication No. US20030036094A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Bolstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James

```

```

APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US/09/904,820
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Sapien
US-09-904-820-236

Query Match          4.1%; Score 7; DB 9; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      8 AAAAAA 14
DB      338 AAAAAA 344

```

```

APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gunney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mathew, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT FILING DATE: US/09/904,859
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Sapien
US-09-904-859-236

Query Match          4.1%; Score 7; DB 9; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      8 AAAAAA 14

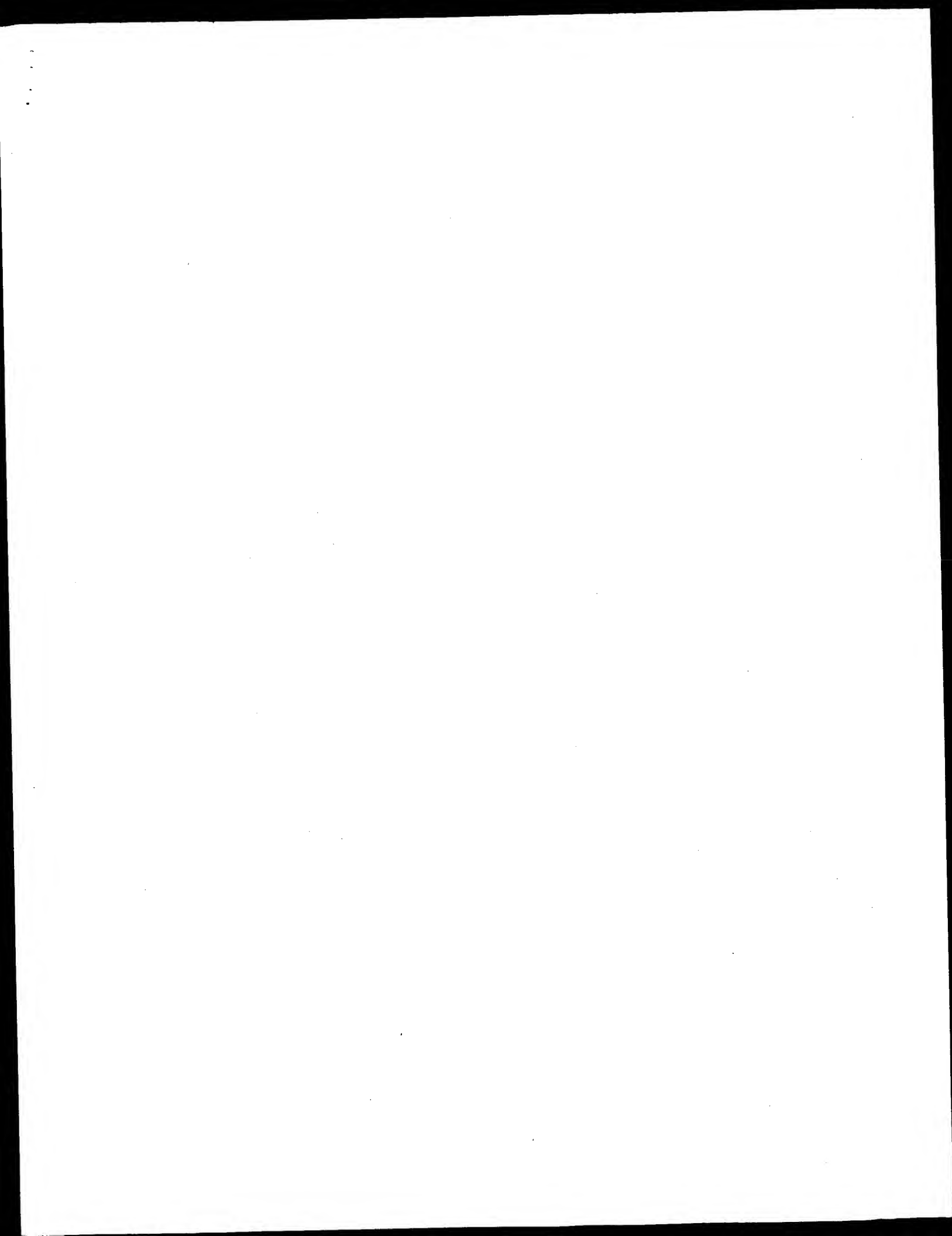
```

Mon Jul 7 08:55:22 2003

us-09-674-779b-2.oli.rapb

Db 338 |||||
AAAAAL 344

Search completed: July 6, 2003, 14:26:44
Job time : 24 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 13:40:40 ; Search time 35 Seconds

(without alignments)
203.827 Million cell updates/sec

Title: US-09-674-779B-2

Perfect score: 861

Sequence: 1 MMHIGTAAALSVITFM.....IAFGTNEAASQNRRAELSY 172

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	316	36.7	166	1	PAL_PSEPU
2	285.5	33.2	173	1	PAL_ECOLI
3	264	30.7	153	1	PAL_HAEIN
4	258	30.0	150	1	PAL_PASMU
5	245	28.5	176	1	PAL_PLEPN
6	185.5	21.5	194	1	OMP_BORAV
7	176.5	20.5	353	1	OMP1_HAEIN
8	174	20.2	326	1	PORF_PSEFL
9	173.5	20.2	353	1	OMP2_HAEIN
10	171.5	19.9	350	1	PORF_PSEAE
11	165.5	19.2	359	1	OMP3_HAEIN
12	163.5	19.0	344	1	PORF_PSESE
13	154	17.9	219	1	YIND_ECOLI
14	151.5	17.6	236	1	OMP3_NEIGO
15	148.5	17.2	243	1	OMP4_ESCEB
16	148.5	17.2	346	1	OMP4_ECOLI
17	148.5	17.2	351	1	OMP4_SHIDY
18	144.5	16.8	242	1	OMP4_NEIMA
19	143	16.6	326	1	YB99_MYCTU
20	141	16.4	238	1	OMP4_CITFR
21	140.5	16.3	350	1	OMP4_SALTY
22	139.5	16.2	243	1	OMP4_ESCHE
23	139	16.1	241	1	OMP4_ESCBL
24	137.5	16.0	350	1	OMP4_ENTAE
25	136.5	15.9	344	1	OMP4_KLEPN
26	135.5	15.7	349	1	OMP4_BUCAI
27	127.5	14.8	293	1	MOTY_VIBAN
28	122.5	14.2	293	1	MOTY_VIBAN
29	122	14.2	359	1	OMP4_SERMA
30	117	13.6	417	1	MOTB_TREPA
31	106.5	12.4	261	1	MOTB_BACSU
32	104	12.1	243	1	OMP4_SEROD
33	100	11.6	218	1	YTXE_BACME

34	98.5	11.4	150	1	YFIB_ECOLI
35	95.5	11.1	238	1	MOTB_TREPA
36	87	10.1	257	1	MOTB_HELPU
37	86	10.0	242	1	YTXE_BACSU
38	83	9.6	551	1	G6P2_STRCO
39	82	9.5	257	1	MOTB_HELPU
40	81	9.4	639	1	COOS_RHORI
41	81	9.4	145	1	PMPE_CITFR
42	80	9.3	361	1	TRMB_CITPN
43	80	9.3	378	1	DPB3_BACSU
44	80	9.3	378	1	SAHH_THEMEA
45	79	9.2	404	1	G6P1_STRCO
46	79	9.2	550	1	HE_PARLI
47	78.5	9.1	587	1	ITR1_PIG
48	78.5	9.1	902	1	ARCQ_HALNI
49	78	9.1	307	1	HMEN_DROVI
50	76	8.8	884	1	ITR3_HUMAN
51	76	8.8	886	1	ITR3_MESAU
52	76	8.8	886	1	ITR3_MOUSE
53	76	8.8	886	1	ITR3_MOUSE
54	75.5	8.8	224	1	VTL2_RAT
55	75.5	8.8	907	1	ITR1_MOUSE
56	75.5	8.8	3649	1	ACYS_NOCIA
57	75	8.7	399	1	TCR1_ECOLI
58	74.5	8.7	279	1	PSN1_PROME
59	74.5	8.7	928	1	PMF9_CHLTP
60	74	8.6	360	1	Y421_METUA
61	74	8.6	360	1	Y105_SULSO
62	74	8.6	887	1	ITR3_RAT
63	73.5	8.5	330	1	IAFU_VIBPA
64	73.5	8.5	928	1	DEOL_ECOLI
65	73.5	8.5	1707	1	194K_TREVS
66	73	8.5	460	1	VATB_METPA
67	73	8.5	535	1	VANI_YEAST
68	72.5	8.4	351	1	COBT_MYCLE
69	72.5	8.4	460	1	PROB_XYLEA
70	72.5	8.4	460	1	VATB_METPA
71	72.5	8.4	521	1	CP11_SPAU
72	72.5	8.4	2431	1	POLN_SFV
73	72	8.4	309	1	MOTB_SALTY
74	72	8.4	320	1	Y054_MYCPN
75	72	8.4	347	1	PYRC_ECO57
76	71.5	8.3	979	1	P115_MYCHR
77	71.5	8.3	1140	1	YW66_YEAST
78	71	8.2	261	1	LPFR_MYCTU
79	71	8.2	347	1	PYRC_ECOLI
80	71	8.2	362	1	PHEA_NEIGO
81	71	8.2	473	1	CYSN_BUCAI
82	71	8.2	540	1	CALX_HELPU
83	71	8.2	606	1	PRIM_MYXXA
84	70.5	8.2	126	1	P15_MOUSE
85	70.5	8.2	416	1	IF5_CAEEL
86	70.5	8.2	441	1	IF5_CAEEL
87	70.5	8.2	552	1	HMEN_DROME
88	70.5	8.2	568	1	DCPY_ZYMO
89	70.5	8.2	839	1	YNEI_YEAST
90	70	8.1	189	1	LPK_MYCTU
91	70	8.1	336	1	USG_PSEAE
92	70	8.1	390	1	DACD_SALTY
93	70	8.1	501	1	YNS7_CONGL
94	69.5	8.1	321	1	YC39_CYAPA
95	69.5	8.1	351	1	HI81_PSEAE
96	69.5	8.1	402	1	DP3B_MYCTU
97	69.5	8.1	805	1	SYFB_MYCPN
98	69.5	8.1	914	1	ITR1_MESAU
99	69	8.0	191	1	UCR1_RHOCA
100	69	8.0	401	1	NIRS_ENTAG

ALIGNMENTS

RESULT 1

PAL_PSEPU STANDARD: PRT: 166 AA.
 ID PAL_PSEPU
 AC PA3036:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL OR PAL1.
 OS Pseudomonas putida.
 OC Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 CC NCBI_TaxID=303;
 CC (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=mt-2;
 RX MEDLINE=96198174; PubMed=8626299;
 RA Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.L.;
 RT "The Pseudomonas putida peptidoglycan-associated outer membrane
 lipoprotein is involved in maintenance of the integrity of the cell
 cell envelope."
 RT J. Bacteriol. 178:1699-1706(1996).
 CC -1- FUNCTION: VERY STRONGLY ASSOCIATED WITH THE PEPTIDOGLYCAN (BY
 SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 anchor.
 CC -1- SIMILARITY: TO OTHER PAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X74218; CAA52294.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE: PS01068; OMPA; 1.
 DR Outer membrane; Signal; Lipoprotein.
 FT SIGNAL 1 21 BY SIMILARITY.
 FT CHAIN 22 166 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
 FT LIPID 22 22 N-ACYL DIGLYCERIDE (BY SIMILARITY).
 FT DOMAIN 98 142 OMPA-LIKE.
 SQ SEQUENCE 166 AA; 17833 MW; 17353181435E5AC1 CRC64;
 Query Match 36.7%; Score 316; DB 1; Length 166;
 Best Local Similarity 43.2%; Pred. No. 9, 8e-22;
 Matches 73; Conservative 34; Mismatches 46; Indels 16; Gaps 6;
 QY 6 QIAAAALSVLTFTMGCAKSTISQ---VMAPNAPPTGYTGYTGYTGAAPVINDETVA 61
 DB 8 KFAALMLAMAV---AVGCCSKGDMCEGAADPNNGYAN---TGA---VDGSISEEA 57
 QY 62 LASKLPLVYEDFSDPEIKQAALIDEQAQFLTTQTAFLVLAAGHTDERGSRREYMSIG 121
 DB 58 -ALRAITTFEFDSSLRKEAMRALDVHAKDKANGN-RVLEGGTDEGTREYMAIG 115
 QY 122 ERRAAVAVRYLILGKINQASVEIISFGERRPIAFGNEEAMSONRAEL 170
 DB 116 ERRRAKAVQRYLVLGVSAPQLELVSYGERPVAATGNDQSMQNRREVL 164

DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL OR EXCC OR B0741 OR Z0909 OR ECS0776.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCBI_TaxID=562, 83334;
 CC (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=87133578; PubMed=3545827;
 RA Chen R., Henning U.;
 RT "Nucleotide sequence of the gene for the peptidoglycan-associated
 lipoprotein of Escherichia coli K12."
 RT Eur. J. Biochem. 163:73-77(1987).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=92244043; PubMed=1574003;
 RA Lazzeroni J.-C., Portalier R.;
 RT "The exc gene of Escherichia coli K-12 required for cell envelope
 integrity encodes the peptidoglycan-associated lipoprotein (PAL).";
 RT Mol. Microbiol. 6:735-742(1992).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RT science 277:1453-1474(1997).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map."
 RT DNA Res. 3:137-155(1996).
 RN [5]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatter F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RT Nature 409:529-533(2001).
 RN [6]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohsudo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Saekawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12."
 RT DNA Res. 8:11-22(2001).
 RN [7]
 RN SEQUENCE OF 1-30 FROM N.A.
 RC STRAIN=JM105;
 RX MEDLINE=90078104; PubMed=2687247;

```

RA Levengood S.K., Webster R.E.;
RT "Nucleotide sequences of the tolA and tolB genes and localization of
RT their products, components of a multistep translocation system in
RT Escherichia coli.";
RL J. Bacteriol. 171:6600-6609(1989).
RN (8)
RP CRYSTALLIZATION.
RX MEDLINE-211173492;
RA Abargel C., Walburger A., Chentivise S., Iazdunski C.;
RT "Crystallization and preliminary crystallographic study of the
RT peptidoglycan-associated lipoprotein from Escherichia coli.";
RL Acta Crystallogr. D 57:317-319(2001).
CC -1- FUNCTION: Thought to play a role in bacterial envelope integrity.
CC Very strongly associated with the peptidoglycan.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -----
CC -1- SIMILARITY: TO OTHER PAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL, X05123; CAA28771.1; -;
DR EMBL, X65796; CAA46673.1; -;
DR EMBL, AE000177; AAC73835.1; -;
DR EMBL, D90713; BAA35407.1; -;
DR EMBL, AE005252; AAG55077.1; -;
DR EMBL, AP002553; BAB34199.1; -;
DR EMBL, M28232; -; NOT_ANNOTATED_CDS.
DR PIR, A27534; LPECPG.
DR PIR, S20547; S20547.
DR ECoGene; EG10684; pal.
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA. 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR PRODOM; PD000930; Bac_OmpA. 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE; PS01068; OMPA. 1.
KW Outer membrane; Signal; Lipoprotein; Complete proteome.
FT CHAIN 1 21
FT LIPID 22 173 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
FT LIPID 22 22 N-ACYL DIGLYCERIDE.
FT DOMAIN 105 149 OMPA-LIKE.
SQ SEQUENCE 173 AA; 18824 MW; 449F959C0274430 CRC64;

Query Match 33.2%; Score 285.5; DB: 1; Length 173;
Best Local Similarity 37.6%; Pred. No.5.9e-19;
Matches 67; Conservative 33; Mismatches 67; Indels 11; Gaps 3;

QY 1 MCHIIQIAAAALSVLTGMTGKANKSTQYVAPNAPTYG---TGVIYGVAPLYVDN 56
DB 1 MQLKVKVLGMLIALPVAIAACSSNKNSN-----DQSEGMLGATGMDANGGNMSS 55
QY 57 ETVAALASKIP-GLVYFEDPSDIKIQAAIILDEQAFITLTNOTARVLAAGHTDERGSR 114
DB 56 EQAFILQMOQLQONNIVTFYDLKDYIRSDFAQMLDAHANFLSNSRYKATVGVGHADERGTP 115
QY 115 EYNNSLIGERRAAVRYNLKGIQNASVEITISFEGRPIAFGTREEMASORRAELSY 172
DB 116 EYNISLGERRANAVKMYLQGGVAGADQISIVSYGKRPVAVLGHDEAVYSKRRRAVLY 173

RESULT 3
PAL_HAEIN STANDARD; PRT; 153 AA.
AC PAL0324;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

DE Outer membrane protein P6 precursor (OMP P6) (15 kDa peptidoglycan-associated lipoprotein) (PC protein).

GN PAL OR OMP6 OR HI0381.

OC Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OX NCBI_TaxID=727;

RP [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=68115138; PubMed=2828309.

RA Deich R.A., Metcalf B.J., Finn C.W., Parley J.E., Green B.A.;

RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated outer membrane lipoprotein and an antigenically related 15,000-dalton protein from Haemophilus influenzae";

RL J. Bacteriol. 170:489-498(1988).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=8608563; PubMed=3257200;

RA Nelson M.B., Apicella M.A., Murphy T.F., Vankeulen H., Spillia L.D.,

RA Rekosh D.;

RT "Cloning and sequencing of Haemophilus influenzae outer membrane protein P6";

RL Infect. Immun. 56:128-134(1988).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Rd / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kienlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Flie L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";

RL Science 269:496-512(1995).

CC -1 SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor.

CC -1 SIMILARITY: TO OTHER PAL PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on ways to use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M19391; AAA24994.1; -

DR EMBL; M18878; AAA24940.1; -

DR EMBL; U32722; AAC22039.1; -

DR PIR; A27558; A27558.

DR PIR; A28543; A28543.

DR TIGR; HI0381. -

DR InterPro: IPR001145; Bac-OmpA.

DR Pfam; PF00691; OmpA. 1.

DR PRINTS; PR01021; OMPADOMAIN.

DR ProDom; PD000930; Bac-OmpA. 1.

DR PROSITE; PS00013; PROKAR_LIPOPROTEIN. 1.

DR PROSITE; PS01068; OMPA. 1.

FT Outer membrane; Signal; Lipoprotein; Complete proteome.

FT SIGNAL 1 19

FT CHAIN 20 153 OUTER MEMBRANE PROTEIN P6.

FT LIPID 20 20 N-ACYL DIGLYCERIDE.

FT DOMAIN 85 129 OMPA-LIKE.

SO SEQUENCE 153 AA; 16108 MW; 3DF358122EF17A11 CRC64;

Query March 30.7%; Score 264; DB 1; Length 153;

Best Local Similarity 49.5%; Pred. NO. 4.4e-17;

Matches 51; Conservative 20; Mismatches 32; Indels 0; Gaps 0;

QY 72 FDPSDEIKPQAAALDEAOFLTTNQTARVLVAGHTDEGRSREYMSIGERRAVRNT 131
 DB 72 FAYDOSTLASKYLPDSVNAQAEYLTHTPGARVMTHGHTDEGRSREYMSIGERRADVAET 131
 QY 132 LIGGICINQASVEIISFGERRPIAFGTNEAMSQNRRAELSY 172
 DB 132 LRMAGVSRQOIRVVSYGKERPANYGHDEASHQNRREVEFTY 172

RESULT 6
 OMPA_BORAV STANDARD; PRT; 194 AA.
 AC Q05146;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1995 (Rel. 30, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Outer membrane protein A precursor.
 GN OMPA.
 OS Bordetella avium.
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OC Bordetella.
 OX NCBI_TaxID=521;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=197;
 RA MEDLINE=93077456; PubMed=1447140;
 RA Gentry-Weeks C.R., Hultsch A.-L., Kelly S.M., Keith J.M.,
 RA Cutliss R. III;
 RA "Cloning and sequencing of a gene encoding a 21-kilodalton outer
 RA membrane protein from Bordetella avium and expression of the gene in
 RA Salmoneella typhimurium.";
 RL J. Bacteriol. 174:7729-7742(1992).
 CC -1- FUNCTION: STRUCTURAL PROTEIN THAT MAY PROTECT THE INTEGRITY OF
 CC THE BACTERIUM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: M96550; AAA22979.1; -
 DR PIR: A45275; A45275.
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; Ompa; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; 1.
 KW Outer membrane; Transmembrane; Porin; Antigen; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 194 OUTER MEMBRANE PROTEIN A.
 FT DOMAIN 54 76 PRO-RICH.
 FT DOMAIN 121 165 OMPA-LIKE.
 SQ SEQUENCE 194 AA; 21115 MW; 1A25FE2A6367DAE7A CRC64;

Query Match 21.5%; Score 185.5; DB 1; Length 194;
 Best Local Similarity 30.7%; Pred. No. 7.3e-10;
 Matches 47; Conservative 22; Mismatches 63; Indels 21; Gaps 2;

QY 18 TFMTCGANKSTQVWAPNAPGTGYTGVAPLVNDDETVALASKLPSLVYFDPDS 77
 DB 56 TGIPGCDGVPAQPEKAPM-----AAKVFNDT-----FEDFS 94

QY 78 EIRPQAAALIDEOAQLTTNQTARVLVAGHTDEGRSREYMSIGERRAVRNTLGGKCI 137
 DB 95 TLKPEGRQLDDVQAQARAIDLETTIAGVNTDSIGTEAYNMKLSERRASVAVYLSKSI 154

QY 138 NQASVELISFGEERPIAFGTNEAMSQNRRAEL 170

DB 155 DPNRITTEKGLNPLTASNKTAGRRANRVEL 187.
 RESULT 7
 OM51_HAEIN STANDARD; PRT; 353 AA.
 AC P43840;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Outer membrane protein P5 precursor (OMP P5).
 GN OMPA OR OMP5 OR H1164.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Gnehm C.L., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
 RA Venter J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RA Rd".
 RL Science 266:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: U32796; AAC2819.1; -
 DR HSSP: P02934; IQDP.
 DR TIGR: H1164; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR InterPro: IPR000498; OMPA_tmem.
 DR Pfam: PF00691; Ompa; 1.
 DR Pfam: PF01389; Ompa_membrane; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; 1.
 KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 353
 FT DISULFID 326 338 BY SIMILARITY.
 FT DOMAIN 272 316 OMPA-LIKE.
 SQ SEQUENCE 353 AA; 37743 MW; 64ACB3E7BFF96B39 CRC64;

Query Match 20.5%; Score 176.5; DB 1; Length 353;
 Best Local Similarity 30.6%; Pred. No. 9.8e-09;
 Matches 48; Conservative 27; Mismatches 61; Indels 21; Gaps 4;

QY 35 PNAPTGY-----TGVAPLVNDDETVALASKLPSLVYFDPDSDEIRPQA 83
 DB 193 PNTALNYPNMGISNAGISYRFGGAAPVAVAPVSKTS-LNSDVTFAFGKANKLKPQA 251

QY 84 AAILDEQAQFLTTNQTARVLVAGHTDEGRSREYMSIGERRAVRNTLGGKICNASVE 143
 DB 252 QATIDSTIGEMSQVSAKAVAVAGYTDRIQSDAFNVKLSQERADSVAVYFAKGVADAIS 311

OY 144 IISGGERPIAFGTNEE-----AMSONRRAELS 171
 DB 312 ATGIGKAMPVTGATCDGVKGRKALLACFADPRRVEIA 348

RESULT 8

PORT_PSEFL STANDARD; PRT; 326 AA.

AC P37726;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane porin F precursor (Root adhesin).

OS OPRF.

OS Pseudomonas fluorescens.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=294;

RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-74 AND 238-258.

RC STRAIN=OE 28.3;

RA MEDLINE=92167971; PubMed=1538702;

RT de Mot R., Proost P., van Damme J., Vanderleyden J.;

RL "Homology of the root adhesin of Pseudomonas fluorescens OE 28.3 with

portin F of P. aeruginosa and P. syringae.";

MOI. Gen. Genet. 231:489-493(1992).

CC -1- FUNCTION: KNOWN TO STABILIZE THE OUTER MEMBRANE.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.

CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO OTHER

CC PSEUDOMONAS SPECIES OPRF.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF115334; AAD45981.1;

DR InterPro: IPR001145; Bac_OmpA.

DR Pfam: PF00691; OmpA.1.

DR PRINTS: PRO1021; OMPADOMAIN.

DR PRODOM: PD000930; Bac_OmpA.1.

DR PROSITE: PS01068; OMPA.1.

KW Outer membrane; Transmembrane; Porin; Signal; Repeat.

FT SIGNAL 1 24

FT CHAIN 25 326 OUTER MEMBRANE PORIN F.

FT DOMAIN 190 205 8 X 2 AA TANDEM REPEATS OF X-P.

FT REPEAT 190 191 1.

FT REPEAT 192 193 2.

FT REPEAT 194 195 3.

FT REPEAT 196 197 4.

FT REPEAT 198 199 5.

FT REPEAT 200 201 6.

FT REPEAT 202 203 7.

FT REPEAT 204 205 8.

FT DOMAIN 253 297 OMPA-LIKE.

SO SEQUENCE 326 AA; 34482 MW; 4C183F56A84D2A76 CRC64;

Query Match 20.2%; Score 174; DB 1; Length 326;
 Best Local Similarity 36.3%; Pred. No. 1.5e-08;
 Matches 37; Conservative 14; Mismatches 51; Indels 0; Gaps 0;

OY 70 VYFPEDSDEIRKPOAAIIDEQAOFLTTNOTARVLVAGHDEKSGREYNSLGERRAVAVR 129

DB 219 VKFPEKSVKPNSTGDKYKNIADPMAQYPAITNEVYAGHTDSIGDPAITNOKLSQRADRVK 278

OY 130 NYLKGGINOASVELISGGERPIAFGTNEEAMSONRRAELS 171

DB 279 QVLVMDGVAPSRITAVGCEGRPAVDANTEAGRAVNRVEAS 320

RESULT 9
 OMS2_HAEIN STANDARD; PRT; 353 AA.

AC P38368;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Outer membrane protein p5 precursor (OMP p5).

GN OMPA OR OMP5.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI_TaxID=727;

RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-34.

RC STRAIN=1613 / Serotype B; 22-34.

RA MEDLINE=93366472; PubMed=8359929;

RT Munson R.S. Jr., Grass S., West R.;

RL "Molecular cloning and sequence of the gene for outer membrane

protein p5 of Haemophilus influenzae.";

Infect. Immun. 61:4017-4020(1993).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.

CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----

DR EMBL: I20309; AAA03346.1;

DR HSSP: P02934; LOUP.

DR InterPro: IPR001145; Bac_OmpA.

DR Pfam: PF00691; OmpA.1.

DR PRINTS: PRO1389; OMPA-membrane.1.

DR PRODOM: PD000930; Bac_OmpA.1.

DR PROSITE: PS01068; OMPA.1.

KW Outer membrane; Transmembrane; Porin; Signal.

FT SIGNAL 1 21

FT CHAIN 22 353 OUTER MEMBRANE PROTEIN P5.

FT DISULFID 326 338 BY SIMILARITY.

FT DOMAIN 272 316 OMPA-LIKE.

SO SEQUENCE 353 AA; 37594 MW; E58A659E7860D0F7 CRC64;

Query Match 20.2%; Score 173.5; DB 1; Length 353;
 Best Local Similarity 30.6%; Pred. No. 1.8e-08;
 Matches 48; Conservative 26; Mismatches 62; Indels 21; Gaps 4;

OY 35 PNAPTYG-----TGVIY---TGVAFLVNDDETVKALASKLPALVYFPEDSDEIRKPOA 83

DB 193 PNTATNYNWMISINAGISYRFGQGAAPVVAPEVYSKTFE-LNSDVFAFGKANKLKPOA 251

OY 84 AALIDEQAOFLTTNOTARVLVAGHDEKSGREYNSLGERRAVAVRNLKGINOASVE 143

DB 252 QATLDSIYCEMSQVSAKAVAGYTDRTGSDPENVKLQERADSVANYFAKVAADAIS 311

Query Match 14.4 IISGGERPIAFGTNEE-----AMSONRRAELS 171
 DB 312 ATGIGKAMPVTGATCDGVKGRKALLACFADPRRVEIA 348

OY 144 IISGGERPIAFGTNEE-----AMSONRRAELS 171

DB 312 ATGIGKAMPVTGATCDGVKGRKALLACFADPRRVEIA 348

OY 130 NYLKGGINOASVELISGGERPIAFGTNEEAMSONRRAELS 171

DB 279 QVLVMDGVAPSRITAVGCEGRPAVDANTEAGRAVNRVEAS 320

OY 130 NYLKGGINOASVELISGGERPIAFGTNEEAMSONRRAELS 171

DB 279 QVLVMDGVAPSRITAVGCEGRPAVDANTEAGRAVNRVEAS 320

OY 130 NYLKGGINOASVELISGGERPIAFGTNEEAMSONRRAELS 171

DB 279 QVLVMDGVAPSRITAVGCEGRPAVDANTEAGRAVNRVEAS 320

OY 130 NYLKGGINOASVELISGGERPIAFGTNEEAMSONRRAELS 171

DB 279 QVLVMDGVAPSRITAVGCEGRPAVDANTEAGRAVNRVEAS 320

OY 130 NYLKGGINOASVELISGGERPIAFGTNEEAMSONRRAELS 171

DB 279 QVLVMDGVAPSRITAVGCEGRPAVDANTEAGRAVNRVEAS 320

OY 130 NYLKGGINOASVELISGGERPIAFGTNEEAMSONRRAELS 171


```

ID      PORE_PSESTY      STANDARD:      PRT:      344 AA.
AC      P22263;
DT      01-AUG-1991 (Rel. 19, Created)
DT      01-AUG-1991 (Rel. 19, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Outer membrane porin F precursor.
GN      OPFR.
OS      Pseudomonas syringae (pv. syringae).
OS      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC      Pseudomonas
CC      NCBI_TaxID=321;
OX      [1]
RN      SEQUENCE FROM N.A.
RP      STRAIN-ATCC 19310;
RX      MEDLINE=91100367; PubMed=1898935;
RA      Ulstrom C.A., Stehnel R., Woodruff W., Steinbach S.,
RA      Hancock R.E.W.;
RT      "Conservation of the gene for outer membrane protein OprF in the
RT      family Pseudomonadaceae: sequence of the Pseudomonas syringae oprF
RT      gene."
RT      J. Bacteriol. 173:768-775(1991).
CC      -1- FUNCTION: HAS PORIN ACTIVITY, FORMING SMALL WATER-FILLED CHANNELS.
CC      ALSO HAS A STRUCTURAL ROLE IN DETERMINING CELL SHAPE AND ABILITY
CC      TO GROW IN LOW-OSMOLARITY MEDIUM.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC      -1- PTM: TWO DISULFIDE BONDS ARE PRESENT.
CC      -1- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO OTHER
CC      PSEUDOMONAS SPECIES OPFR.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL: M55408; AAA25910.1; -.
DR      PIR: A39139; A39139.
DR      InterPro: IPR001145; Bac_OmpA.
DR      Pfam: PF00691; OmpA; 1.
DR      PRINTS: PR01021; OMPADOMAIN.
DR      ProDom: PD000930; Bac_OmpA; 1.
DR      PROSITE: PS01068; OMPA; 1.
KW      Outer membrane; Transmembrane; Porin; Signal.
FT      SIGNAL          1      24      BY SIMILARITY.
FT      CHAIN           25      344      OUTER MEMBRANE PORIN F.
FT      DOMAIN          270      315      OMPA-LIKE.
SQ      SEQUENCE      344 AA; 36567 MW; 5B463D9AD973D755 CRC64;

Query Match      19.0%; Score 163.5; DB: 1; Length 344;
Best Local Similarity 35.6%; Pred. No. 1.4e-07;
Matches 36; Conservative 14; Mismatches 50; Indels 1; Gaps 1;

OY      70 VFPEPDSDEIKPQAAALIDEQAOFLTNQTRARVLVAGHTDERSREYVMSLGERRAAVVR 129
      1 1111 1111 1111 1111 1111 1111 1111 1111 1111
DB      236 VKFPEKDSVSVKPNYSVGDIKLNADPMQOYPTTTVEGHGTSVGDATNOKRKERRANAVK 295
      1 1111 1111 1111 1111 1111 1111 1111 1111 1111

OY      130 NYLLGK-GINQASVEIISFGGERPIAGTNEANSQNRRAE 169
      1 1111 1111 1111 1111 1111 1111 1111 1111 1111
DB      296 QVLVNOYGVGASRVNSVGYGSKPEVADNATAGRAVNRKVE 336
      1 1111 1111 1111 1111 1111 1111 1111 1111 1111

RESULT 13
YIAD_ECOLI
ID      YIAD_ECOLI      STANDARD:      PRT:      219 AA.
AC      P37665;
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical lipoprotein yIad precursor.
DE      YIAD OR B3552.
OS      Escherichia coli.

```

ID	OMP3_NEIGO	STANDARD;	PRT;	236 AA.
AC	P07050;			
CT	01-APR-1988 (Rel. 07, Created)			
CC	Bacteria;Proteobacteria; gamma subdivision; Enterobacteriaceae;			
CC	Escherichia			
CC	NCBI_TaxID=562;			
CC	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-K12 / MG1655;			
RC	MEDLINE=94316500; PubMed=8041620;			
RA	Sofia H.J., Burland V., Daniels D.L., Plunkett G. II, Blattner F.R.;			
RA	"Analysis of the Escherichia coli genome. V. DNA sequence of the			
RT	region from 76.0 to 81.5 minutes."			
RL	Nucleic Acids Res. 22:2576-2586(1994).			
RN	[2]			
RP	SEQUENCE OF 1-104 FROM N.A.			
RX	MEDLINE=90202748; PubMed=18180922;			
RA	Pierison D.E., Campbell A.;			
RT	"Cloning and nucleotide sequence of bise, the structural gene for			
RT	biotin sulfoxide reductase in Escherichia coli."			
RL	J. Bacteriol. 172:2194-2198(1990).			
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor			
CC	(Probable).			
CC	-1- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO PSEUDOMONAS			
CC	OPR.			
CC	-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A			
CC	FRAMESHIFT IN POSITION 88.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U000039; AAC18529.1; -			
DR	EMBL; AE000432; AAC76576.1; -			
DR	EMBL; M34827; -; NOT ANNOTATED_CDS.			
DR	Ecogene; EG12271; yfad.			
DR	InterPro; IPR001145; Bac_OmpA.			
DR	Pfam; PF00691; OmpA; 1.			
DR	PRINTS; PRO1021; OMPADOMAIN.			
DR	ProDom; PD000930; Bac_OmpA; 1.			
DR	PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.			
DR	PROSITE; PS01068; OMPA; 1;			
KW	Hypothetical protein; Membrane; Lipoprotein; Signal;			
KW	Complete proteome.			
FT	SIGNAL	20	POTENTIAL.	
FT	CHAIN	1		
FT	LIPID	21	219	HYPOTHETICAL LIPOPROTEIN YFAD.
FT	DOMAIN	148	192	N-ACYL DIGLYCERIDE (POTENTIAL).
SO	SEQUENCE	219 AA;	22169 MW;	8BDEDE55A0A6EE368 CRC64;
Query Match		17.98;	Score 154;	DB 1; Length 219;
Best Local Similarity		31.28;	Pred. No. 6e-07;	
Matches	39;	Conservative	21;	Mismatches 61; Indels 4; Gaps 1;
OY	47	TGVAPELVNDNETVYALASKLPSELYVYFPEDSDIEIKPQAAALILDEQAFITNQTQARVYLVAG	106	
Db	95	TGVSVTSSGDNIT-----LNMFPNNVTFPSSSRTLKPGAGANTLTGVAMLYKREYKPTAVNYIG	150	
OY	107	HTDEGSGREXNMSLGERRAAVARNYLLGKGINGASVEIIISFGEEPRPIAFGTNEBAMSONR	166	
Db	151	YTDSTGCHDLMRLRISQGRADSVASALITGGVDASRIQTGLGPNPFIASNSTAEKGQRNR	210	
OY	167	RAELS 171		
Db	211	RVEIT 215		
RESULT 14				
ID	OMP3_NEIGO	STANDARD;	PRT;	236 AA.
AC	P07050;			
CT	01-APR-1988 (Rel. 07, Created)			

DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Outer membrane protein P.III precursor (Gonococcal protein III)
DE (P.III).
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87139801; PubMed=3102671;
RA Gotschlich E.C., Seiff M., Blake M.S.;
RT "The DNA sequence of the structural gene of gonococcal protein III and the flanking region containing a repetitive sequence. Homology of protein III with enterobacterial Ompa proteins.";
RL J. Exp. Med. 165:471-482(1987).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- MISCELLANEOUS: P.III IS CLOSELY ASSOCIATED WITH P.II, BUT NOT ALL OF THE P.II MOLECULES IN THE GONOCOCCAL OUTER MEMBRANE.
CC -1- MISCELLANEOUS: A PORTION OF THE P.III IS EXPOSED TO THE SURFACE IN INTACT GONOCOCCI; THE PROTEIN CAN BE LABELED WITH IODINE AND REACTS WITH MABS.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO N.MENINGITIDIS RMPM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05105; CAA28752.1; -;
DR PIR: A27894; A27894.
DR InterPro: IPR001145; Bac_Ompa.
DR Pfam: PF00691; Ompa.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_Ompa; 1.
DR PROSITE: PS01068; Ompa; 1.
KW Outer membrane; Porin; Transmembrane; Antigen; Signal; Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 236 OUTER MEMBRANE PROTEIN P.III.
FT DOMAIN 69 76 4 X 2 AA TANDEM REPEATS OF X-P.
FT REPEAT 69 70 1.
FT REPEAT 71 72 2.
FT REPEAT 73 74 3.
FT REPEAT 75 76 4.
FT DOMAIN 131 175 OMPA-LIKE.
FT DISULFID 185 208 BY SIMILARITY.
SQ SEQUENCE 236 AA; 25540 MM; DAE9AAECA6FB199 CRC64;
Query Match 17.6%; Score 151.5; DB 1; Length 236;
Best Local Similarity 36.5%; Pred. No. 1.1e-06;
Matches 35; Conservative 13; Mismatches 43; Indels 5; Gaps 1;
QY 56 DETVATLASKLPSTLYVEFDPDSDEIKPQAAALIDEOAFLTNGARVAVAGHTDEQRSRE 115
DB 88 DETISLSAKTL-----FGFDKDSLRAEQDNLKYLQRLSTNTVQSRVSEHTDPMSEK 142
QY 116 YNMSLGERRAVAVRNYYLLGKGINOASVEIISFGEER 151
DB 143 YNOALSERRAYVANNIVNGVPASRISAVGLGESQ 178
RESULT 15
OMPA_ESCFE STANDARD; PRT; 243 AA.
AC P24747;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Outer membrane protein A (Outer membrane protein II) (Fragment).
GN OMPA.

OS Escherichia fergusonii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=564;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35471, ATCC 35469, and ATCC 35472;
RX MEDLINE=92065252; PubMed=1955870;
RA Lawrence J.G., Ochman H., Hartl D.L.;
RT "Molecular and evolutionary relationships among enteric bacteria.";
RL J. Gen. Microbiol. 137:1911-1921(1991).
CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHASES. ALSO ACTS AS A PORIN WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL SOLUTES (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAINS ATCC 35471 AND ATCC 35472.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63352; AAA24236.1; -;
DR EMBL: M63351; AAA24232.1; -;
DR HSSP: M63353; AAA24240.1; -;
DR HSSP: P02934; 10JP.
DR InterPro: IPR001145; Bac_Ompa.
DR InterPro: IPR000498; Ompa_tmem.
DR Pfam: PF00691; Ompa; 1.
DR Pfam: PF01389; Ompa_membrane; 1.
DR ProDom: PD000930; Bac_Ompa; 1.
DR PROSITE: PS01068; Ompa; 1.
KW Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.
FT NON_TER 1 1
FT TRANSMEM 1 8 POTENTIAL.
FT TRANSMEM 12 27 POTENTIAL.
FT TRANSMEM 48 59 POTENTIAL.
FT TRANSMEM 65 81 POTENTIAL.
FT TRANSMEM 87 98 POTENTIAL.
FT DOMAIN 107 114 4 X 2 AA TANDEM REPEATS OF A-P.
FT REPEAT 107 108 1.
FT REPEAT 109 110 2.
FT REPEAT 111 112 3.
FT REPEAT 113 114 4.
FT DOMAIN 163 207 OMPA-LIKE.
FT DISULFID 217 229 BY SIMILARITY.
FT VARIANT 39 39 E -> D (IN STRAIN ATCC 35469).
FT NON_TER 243 243
SQ SEQUENCE 243 AA; 26144 MM; B89F7FC29D42E3AB CRC64;
Query Match 17.2%; Score 148.5; DB 1; Length 243;
Best Local Similarity 31.6%; Pred. No. 2.1e-06;
Matches 42; Conservative 20; Mismatches 60; Indels 11; Gaps 2;
QY 49 VAPLVNDEYKALAKLPSTLYVEFDPDSDEIKPQAAALIDEOAFLTNGARVAVAGHTDEQRSRE 106
DB 106 VAPAPAPAEVQKHTLTLSVDVLFNFKATLPEGAALDQXLSQLSNIDPKDGSVYVLG 165
QY 107 HTDEGSRREYNMSLGERRAVAVRNYYLLGKGINOASVEIISFGEERPIAGTNE----- 159
DB 166 YTRIGSDAVNGQLSERRAQSVYDYLISGIPADKISAGMGESNVTGTCDDNNKQRAA 225
QY 160 --EAMSONRAEL 170
DB 226 LIDCLAPDRRVEI 238

RESULT 16
OMPA_ECOLI STANDARD: PRT: 346 AA.
AC P02934;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (Outer membrane protein II*).
GN OMPA OR TOLG OR TUP OR CON OR B0957 OR Z1307 OR ECS1041.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MEDLINE=81053729; PubMed=6253901;
RT Beck E., Bremer E.;
RL "Nucleotide sequence of the gene ompa coding the outer membrane protein II of Escherichia coli K-12.";
RN [2]
RP Nucleic Acids Res. 8:3011-3024(1980).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MEDLINE=81170587; PubMed=6260961;
RT Movva N.R., Nakamura K., Inouye M.;
RL "Gene structure of the ompa protein, a major surface protein of Escherichia coli required for cell-cell interaction.";
RN [4]
RP J. Mol. Biol. 143:317-328(1980).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RT Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12.";
RN [6]
RP Science 277:1453-1474(1997).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MEDLINE=97061202; PubMed=8905232;
RT Oshima T., Aliba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sampei H., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horinouchi T.;
RL "A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";
RN [8]
RP DNA Res. 3:137-155(1996).
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RA MEDLINE=21074935; PubMed=11206551;
RT Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattnerman T.S.;
RL "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
RN [10]
RP Nature 409:529-533(2001).
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RA MEDLINE=21156231; PubMed=11258796;
RT Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakajima K., Murata T., Tanaka M., Tode T.,

RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T., Kuwara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [12]
RP SEQUENCE OF 22-346.
RC STRAIN-K12;
RA MEDLINE=81054820; PubMed=7001461;
RT Chen R., Schindlmyr W., Kramer C., Chen-Schmeissner U., Henning U.;
RL "Primary structure of major outer membrane protein II (ompa protein) of Escherichia coli K-12.";
RN [13]
RP Proc. Natl. Acad. Sci. U.S.A. 77:4592-4596(1980).
RN [14]
RP SEQUENCE OF 22-32.
RC STRAIN-K12 / W3110;
RA Pasquall C., Sanchez J.-C., Ravlier F., Golaz O., Hughes G.J., Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A., Hochstrasser D.F.;
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
RN [15]
RP SEQUENCE OF 22-34.
RC STRAIN-K12 / EMG2;
RA MEDLINE=97443975; PubMed=9298646;
RT Link A.J., Robison K., Church G.M.;
RL "Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12.";
RN [16]
RP Electrophoresis 18:1259-1313(1997).
RN [17]
RP SEQUENCE OF 22-26.
RC STRAIN-K12 / W3110;
RA MEDLINE=98291876; PubMed=9629924;
RT Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traut M., Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;
RL "Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis.";
RN [18]
RP MUTANTS RESISTANT TO PHAGE ENTRY.
RA MEDLINE=84264337; PubMed=6086577;
RT Morona R., Klose M., Henning U.;
RL "Escherichia coli K-12 outer membrane protein (Ompa) as a bacteriophage receptor: analysis of mutant genes expressing altered proteins.";
RN [19]
RP J. Bacteriol. 159:570-578(1984).
RN [20]
RP MUTANTS RESISTANT TO PHAGE ENTRY.
RA MEDLINE=86033606; PubMed=3902787;
RT Morona R., Kramer C., Henning U.;
RL "Bacteriophage receptor area of outer membrane protein Ompa of Escherichia coli K-12.";
RN [21]
RP J. Bacteriol. 164:539-543(1985).
RN [22]
RP PORIN ACTIVITY.
RC STRAIN-K12;
RA MEDLINE=92129334; PubMed=1370823;
RT Sugawara E., Nikaide H.;
RL "Pore-forming activity of Ompa protein of Escherichia coli.";
RN [23]
RP J. Biol. Chem. 267:2507-2511(1992).
RN [24]
RP TOPOLOGY.
RA MEDLINE=94148615; PubMed=8106193;
RT Gromiha M.M., Ponnuswamy P.K.;
RL "Prediction of transmembrane beta-strands from hydrophobic characteristics of proteins.";
RN [25]
RP Int. J. Pept. Protein Res. 42:420-431(1993).
RN [26]
RP TOPOLOGY.
RA MEDLINE=99296577; PubMed=10368142;
RT Koebnik R.;
RL "Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein Ompa from Escherichia coli.";
RN [27]
RP J. Bacteriol. 181:3688-3694(1999).

RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bisai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 273101; CAA97374.1; -
 DR EMBL: AE006979; AAK45169.1; -
 DR TIGR: MT0922; -
 DR Tuberculist; Rv0899; -
 DR InterPro; IPR001145; Bac_OmpA.
 DR Pfam; PF00691; OmpA; 1.
 DR PRINTS; PRO1021; OMPADOMAIN.
 DR ProDom; PD000930; Bac_OmpA; 1.
 DR PROSITE; PS01068; OMPA; FALSE_NEG.
 KW Hypothetical protein; transmembrane; Complete proteome.
 FT TRANSMEM 30 50
 FT TRANSMEM 101 121 POTENTIAL.
 FT SEQUENCE 326 AA; 33574 MW; 9E0D46ABCC179F4A CRC64;
 SQ
 Query Match 16.6%; Score 143; DB 1; Length 326;
 Best local Similarity 23.3%; Pred. No. 9.5e-06;
 Matches 52; Conservative 26; Mismatches 85; Indels 60; Gaps 4;
 QY 8 AAAAAALSVLTFTGCAKSTSOVMAP-----NATPGTGTG-----V 44
 DB 101 AAKAALMTALNGLIADGVNVDQIHVDPVRLDSSAPVFAVPIPDGLKVERDTV 160
 QY 45 IYGVAPLVNDDETVK-ALASKLPSTL----- 69
 DB 161 TLGTAPSSSEHKAVRAATSTWPMKTYNNITEVGQAPRPPASPCADLSATNAVYTG 220
 QY 70 --VYFDPDSDEIKPQAAAILDEQAQFLTTNQTARVLVAGHTDERGSRXYNMSLGERRAYA 127
 DB 221 GPITAFGNDGASLIPADYEILNRVADKLKACPDARVTINGTDTWTGSEGINIPLSAQRAKI 280
 QY 128 VRRVLLGKGINQASVELISFGERRPIAFGNEAMSONRAEL 170
 DB 281 VADYLVARGVAGDHIAITVGISVNPATSNATPEGRAKNRREVI 323
 RESULT 20
 OMPA_CITR STANDARD; PRT; 238 AA.
 AC P24016;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Outer membrane protein A (Outer membrane protein II) (Fragment).
 GN OMPA.
 OS Citrobacter freundii.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Citrobacter.
 CC NCBI_Taxid=546;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OS60;

RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria";
 RL J. Gen. Microbiol. 137:1911-1921(1991)
 CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF WAITING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTIONS (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M63354; AAA23095.1; -
 DR HSSP; P02934; 10UP
 DR InterPro; IPR001145; Bac_OmpA.
 DR InterPro; IPR000498; OmpA_tmem.
 DR Pfam; PF01389; OmpA_membrane; 1.
 DR ProDom; PD000930; Bac_OmpA; 1.
 DR PROSITE; PS01068; OMPA; 1.
 KW Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.
 FT NON_TER 1 8
 FT TRANSMEM 1 8 POTENTIAL.
 FT TRANSMEM 12 27 POTENTIAL.
 FT TRANSMEM 43 54 POTENTIAL.
 FT TRANSMEM 60 76 POTENTIAL.
 FT TRANSMEM 82 93 POTENTIAL.
 FT DOMAIN 104 109 3 X 2 AA TANDEM REPEATS OF A-P.
 FT REPEAT 104 105 1.
 FT REPEAT 106 107 2.
 FT REPEAT 108 109 3.
 FT DOMAIN 158 202
 FT DISULFID 212 224 OMPA-LIKE.
 FT NON_TER 238 238 BY SIMILARITY.
 SQ SEQUENCE 238 AA; 25664 MW; 66EA136D73A3F286 CRC64;
 Query Match 16.4%; Score 141; DB 1; Length 238;
 Best local Similarity 28.3%; Pred. No. 9.9e-06;
 Matches 43; Conservative 24; Mismatches 61; Indels 24; Gaps 4;
 QY 43 GVIYGV-----APLVNDETVKALASK--LPSLYFDPDSDEIKPQAAAILDE 89
 DB 82 GLISGVSYRFGQGEAPVAVAPAPAEVQTKHFTLSKDSVLFNFNKATLKEPQOALDQ 141
 QY 90 QAQFLTT--NOTARVLVAGHTDERGSRXYNMSLGERRAYVNRNLLGKGINQASVELISF 147
 DB 142 MYSQLSNLDPKDGVVVLGFTDRIGSDAYNGLSEKRSYVVDYLISKIPSKISARGM 201
 QY 148 GEERPIAFGTNE-----EAMSONRAEL 170
 DB 202 GESNPVTGNTCDNVKARAALLDCLAPDRRVEL 233
 RESULT 21
 OMPA_SALTY STANDARD; PRT; 350 AA.
 AC P02936;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Outer membrane protein A precursor (Outer membrane protein 33K) (Outer
 DE membrane major heat-modifiable protein).
 GN OMPA OR STM1070.
 OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 CX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83287368; PubMed=6349993;
 RA Freudi R., Cole S.T.;
 RT "Cloning and molecular characterization of the ompa gene from
 RT Salmonella typhimurium";
 RL Eur. J. Biochem. 134:497-502(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lattelle P.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2";
 RL Nature 413:852-856(2001).
 CC -I- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC -I- SUBUNIT: MONOMER (PROBABLE).
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -I- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X02006; CA26037.1; -
 CC DR EMBL; AE008746; AAL20003.1; -
 CC DR PIR; A03436; MMEBAT.
 CC DR HSSP; P02934; LOUP.
 CC Styene; SG10263; OMPA.
 CC InterPro; IPR001145; Bac_OmpA.
 CC DR InterPro; IPR000498; OmpA_tmem.
 CC Pfam; PF00691; OmpA; 1.
 CC DR ProDom; PD000930; Bac_OmpA; 1.
 CC DR ProSITE; PS01068; OMPA; 1.
 CC Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
 KW Portin; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 1 350
 FT TRANSMEM 22 40 OUTER MEMBRANE PROTEIN A.
 FT TRANSMEM 55 67 POTENTIAL.
 FT TRANSMEM 70 85 POTENTIAL.
 FT TRANSMEM 97 107 POTENTIAL.
 FT TRANSMEM 111 126 POTENTIAL.
 FT TRANSMEM 146 157 POTENTIAL.
 FT TRANSMEM 163 179 POTENTIAL.
 FT TRANSMEM 185 196 POTENTIAL.
 FT DOMAIN 205 212 4 X 2 AA TANDEM REPEATS OF A-P.
 FT REPEAT 205 212 1.
 FT REPEAT 207 206 2.
 FT REPEAT 209 210 3.
 FT REPEAT 211 212 4.
 FT DOMAIN 261 305 OMPA-LIKE.
 FT DISULFID 315 327 BY SIMILARITY.
 FT CONFLICT 114 114 V -> F (IN REF. 1).
 FT CONFLICT 247 247 S -> I (IN REF. 1).
 SQ SEQUENCE 350 AA; 37515 MW; BAAC52C8C5D5E54FE CMC64;

Query Match 16.3%; Score 140.5; DB 1; Length 350;
 Best Local Similarity 30.1%; Pred. No. 1.7e-05;
 Matches 40; Conservative 20; Mismatches 62; Indels 11; Gaps 2;
 OY 49 VAPLVNDNEVYKALSKLPILVYFDPDSDEIKPQAAAILDQAOFLTT--NQTARYVAG 106
 DB 204 VAPAPAPPEVQYHTFTLSKYLVLENKSTLKPGQALDQVLSQSLNDPKDGSVVVG 263
 OY 107 HTDERGSKREYMSLGERRAVAVRYLLGKINGASVEIISFGEERPIAFGTNE----- 159
 DB 264 FTDRIGSDAVYNGSLSEKRAQSVVDYLISKIPSPKISARGGSENPYGTWCDVVRPA 323
 OY 160 --EAMSONRRREL 170
 DB 324 LIDLIADPRVEI 336
 RESULT 22
 OMPA_ESCHE
 ID OMPA_ESCHE STANDARD: PRT: 243 AA.
 AC P24734;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Outer membrane protein A (Outer membrane protein II) (Fragment).
 GN OMPA.
 OS Escherichia hermannii.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OX Escherichia.
 CX NCBI_TaxID=565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33650, and ATCC 33652;
 RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria";
 RL J. Gen. Microbiol. 137:1911-1921(1991).
 CC -I- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC -I- SUBUNIT: MONOMER (PROBABLE).
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -I- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M63346; AAA24234.1; -
 CC DR EMBL; M63347; AAA24238.1; -
 CC DR HSSP; P02934; IBXW.
 CC DR InterPro; IPR001145; Bac_OmpA.
 CC DR InterPro; IPR000498; OmpA_tmem.
 CC Pfam; PF00691; OmpA; 1.
 CC DR ProDom; PD000930; Bac_OmpA; 1.
 CC DR ProSITE; PS01068; OMPA; 1.
 CC Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.
 KW NON_TER 1 1
 FT TRANSMEM <1 8 POTENTIAL.
 FT TRANSMEM 12 27 POTENTIAL.
 FT TRANSMEM 47 58 POTENTIAL.
 FT TRANSMEM 64 80 POTENTIAL.
 FT TRANSMEM 86 97 POTENTIAL.
 FT DOMAIN 107 114 4 X 2 AA TANDEM REPEATS OF A-P.
 FT REPEAT 107 108 1.
 FT REPEAT 109 110 2.

FT REPEAT 111 112 3.
 FT REPEAT 113 114 4.
 FT DOMAIN 163 207 OMPA-LIKE.
 FT DISULFID 217 229 BY SIMILARITY.
 FT NON_TER 243 243
 SQ SEQUENCE 243 AA; 26202 MW; AAGCB6BA390D1E9D CRC64;

Query Match 16.2%; Score 139.5; DB 1; Length 243;
 Best Local Similarity 30.1%; Pred. No. 1.4e-05;
 Matches 40; Conservative 20; Mismatches 62; Indels 11; Gaps 2;

49 VAPLVNDFTVKALSPDYFDPDSDEIKPOAAAIIDEOAQFLTT--NOTARVLVAG 106
 106 VAPAPAPREVQKHTLKSVDYLFNFKATLKPEGGQALDQWYQLSNLDPKDGSVVLG 165
 107 HTDGRSREYNMSLGERRAVAVNYLLGKINQASVEIISFGEERPIAFGTYE----- 159
 166 FTDRIGSDAYNQLSEKRAQSVVDYLISKIPSDKISANGMGESNFTGTCDNVKPRAA 225
 160 --EAMSONRRAEL 170
 226 LIDCLAPDRRVEI 238

RESULT 23

OMPA_ESCBL STANDARD; PRT: 241 AA.

AC Q99124; Q03617;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Outer membrane protein A (Outer membrane protein II) (Fragment).
 GN OMPA.
 OS Escherichia blattae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=563;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33429, and ATCC 33430;
 RX MEDLINE=9206525; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria";
 J. Gen. Microbiol. 137:1911-1921(1991).
 CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 SOLUTES (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATCC 33429.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M63343; AAA24235.1; -;
 DR EMBL: M63344; AAA24239.1; -;
 DR EMBL: M63345; AAA24242.1; -;
 DR HSSP: P02934; IQP.
 DR InterPro: IPR001145; Bac_OmpA.
 DR InterPro: IPR000498; OmpA_tmam.
 DR Pfam: PF00691; OmpA.1.
 DR Pfam: PF01389; OmpA_membrane.1.
 DR ProDom: PD000930; Bac_OmpA.1.
 DR ProSITE: PS01068; OmpA.1.
 KW Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.

FT NON_TER 1 1
 FT TRANSMEM <1 8 POTENTIAL.
 FT TRANSMEM 12 27 POTENTIAL.
 FT TRANSMEM 46 57 POTENTIAL.
 FT TRANSMEM 63 79 POTENTIAL.
 FT TRANSMEM 85 96 POTENTIAL.
 FT DOMAIN 105 112 4 X 2 AA TANDEM REPEATS OF A-P.
 FT REPEAT 105 106 1.
 FT REPEAT 107 108 2.
 FT REPEAT 109 110 3.
 FT REPEAT 111 112 4.
 FT DOMAIN 161 205 OMPA-LIKE.
 FT DISULFID 215 227 V -> G (IN STRAIN ATCC 29907).
 FT VARIANT 33 33 W -> M (IN STRAIN ATCC 33430).
 FT VARIANT 56 56 MISSING (IN STRAIN ATCC 33430).
 FT VARIANT 101 101 N -> S (IN STRAIN ATCC 29907 AND
 STRAIN ATCC 33430).
 FT VARIANT 213 213 STRAIN ATCC 29907 AND
 STRAIN ATCC 33430).
 FT VARIANT 220 220 STRAIN ATCC 33430).
 SQ SEQUENCE 241 AA; 25986 MW; 7C81CC0CCD5F1C40 CRC64;

Query Match 16.1%; Score 139; DB 1; Length 241;
 Best Local Similarity 27.0%; Pred. No. 1.5e-05;
 Matches 41; Conservative 25; Mismatches 62; Indels 24; Gaps 3;

43 GVIYGVAPLVNDFTVKALAS-----KLPSLYFPDPDSDEIKPOAAAIIDE 89
 85 GMLSVGSYFRGQDAPVAPAPAPAPQVOTKHTLSVDYLFNFKSTLKEEGQALDQ 144
 90 QAOFLTT--NOTARVLVAGHTDGRSREYNMSLGERRAVAVNYLLGKINQASVEIISF 147
 145 LYTQLSNLDPKDGAVNYLLGYTRIGSDAYNQLSQGRASVVDYLVSGIPAKITAGQ 204
 148 GEERPIAFGTYE-----EAMSONRRAEL 170
 205 GESNFTGTCDNVKRAALIDCLAPDRRVEI 236

RESULT 24

OMPA_ENTAE STANDARD; PRT: 350 AA.

AC P09146;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Outer membrane protein A precursor.
 GN OMPA.
 OS Enterobacter aerogenes (Aerobacter aerogenes).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Enterobacter.
 OX NCBI_Taxid=548;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84108348; PubMed=6363059;
 RA Braun G., Cole S.T.;
 RT "Molecular characterization of the gene coding for major outer
 membrane protein OmpA from Enterobacter aerogenes";
 Eur. J. Biochem. 137:495-500(1983).
 CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 SOLUTES (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X00254; CAA25062.1; -
 DR PIR: S07222; S07222.
 DR HSSP: P02934; 103P.
 DR InterPro: IPR001145; Bac_OmpA.
 DR InterPro: IPR000498; OmpA_tmam.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01389; OmpA_membrane; 1.
 DR PRODOM: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OmpA; 1.
 KW Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
 KW Porin.
 FT SIGNAL 1 21
 FT CHAIN 22 350
 FT TRANSMEM 27 40
 FT TRANSMEM 59 71
 FT TRANSMEM 74 89
 FT TRANSMEM 101 111
 FT TRANSMEM 115 130
 FT TRANSMEM 146 157
 FT TRANSMEM 163 179
 FT TRANSMEM 185 196
 FT DOMAIN 205 212
 FT REPEAT 205 206
 FT REPEAT 207 208
 FT REPEAT 209 210
 FT REPEAT 211 212
 FT DOMAIN 211 305
 FT DISULFID 315 327
 SQ SEQUENCE 350 AA; 37575 MW; 6276C6F2F21065DA CRC64;
 Query Match 16.0%; Score 137.5; DB 1; Length 350;
 Best local Similarity 31.5%; Pred. No. 3.2e-05;
 Matches 39; Conservative 16; Mismatches 58; Indels 11; Gaps 2;

QY 36 MAPTGYITGVYITGVAVLVNDFTVYALASKLPISLYVDFDSDEIRPOAAALIDEOAOFLT 95
 DB 200 MAPV-----VAPAPAPAEVITKTFTLKSDVLFENNKATLKPEGOALDOLYTQLS 250
 QY 96 T--NQTARVLVAGHTDEGRSREYNMSIGERRAVAVRNYTLGGINQASVEIISFEERPI 153
 DB 251 NMDPKGSAVVLGYIDRISSEQYNOKLSKRAQSVVDILVAKGIDPANKISARGMSDPV 310
 QY 154 AFGT 157
 DB 311 TGNT 314

RESULT 25
 OMPA_KLEPN STANDARD; PRT; 344 AA.
 ID OMPA_KLEPN P24017; 069435;
 DT 01-MAR-1992 (Rel. 21; Created)
 DT 15-DEC-1998 (Rel. 37; Last sequence update)
 DT 30-MAY-2000 (Rel. 39; Last annotation update)
 DE Outer membrane protein A precursor (Outer membrane protein II).
 GN OMPA.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RV 308;
 RX MEDLINE=98192544; PubMed=9524233;
 RA Nguyen T.N., Samuelson P., Sterky F., Merle-Polite C., Robert A.,
 RA Baussant T., Haeuw J.F., Uhlen M., Binz H., Stahl S.;
 RT "Chromosomal sequencing using a PCR-based biotin-capture method
 RT allowed isolation of the complete gene for the outer membrane protein

RT A of Klebsiella pneumoniae.";
 RL Gene 210:93-101(1998).
 RN [2]
 RP SEQUENCE OF 93-335 FROM N.A.
 RC STRAIN=LD119;
 RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria.";
 RL J. Gen. Microbiol. 137:1911-1921(1991).
 CC - FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC - SUBUNIT: MONOMER (PROBABLE).
 CC - SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC - SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ000998; CAA04450.1; -
 DR EMBL: M63355; AAB25119.1; -
 DR HSSP: P02934; 103P.
 DR InterPro: IPR001145; Bac_OmpA.
 DR InterPro: IPR000498; OmpA_tmam.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OmpA; 1.
 KW Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
 KW Porin.
 FT SIGNAL 1 2
 FT CHAIN 15 344
 FT TRANSMEM 15 28
 FT TRANSMEM 48 60
 FT TRANSMEM 63 78
 FT TRANSMEM 90 100
 FT TRANSMEM 104 119
 FT TRANSMEM 140 151
 FT TRANSMEM 157 173
 FT TRANSMEM 179 190
 FT DOMAIN 199 206
 FT REPEAT 199 200
 FT REPEAT 201 202
 FT REPEAT 203 204
 FT REPEAT 205 206
 FT REPEAT 255 299
 FT DOMAIN 309 321
 FT DISULFID 309 321
 FT CONFLICT 335 335
 SQ SEQUENCE 344 AA; 37061 MW; AC88AAE3B7871B16 CRC64;
 Query Match 15.9%; Score 136.5; DB 1; Length 344;
 Best local Similarity 30.2%; Pred. No. 3.9e-05;
 Matches 35; Conservative 20; Mismatches 50; Indels 11; Gaps 2;

QY 66 LPSLYVDFDSDEIRPOAAALIDEOAOFLT--NQTARVLVAGHTDEGRSREYNMSIGER 123
 DB 215 LKSDVLFENFKATLKPEGOALDOLYTQLSNMDPKGSAVVLGYIDRISSEAYNOLSEK 274
 QY 124 RAVAVRNYTLGGINQASVEIISFEERPIANGTNE-----EAMSONRAEL 170
 DB 275 RAQSVVDILVAKGIDPANKISARGMSNPNVTGNTCDNVKARAALIDCLAPDRIVEI 330

RESULT 26
 OMPA_BUCAL

ID OMPA_BUCAT STANDARD: PRT; 349 AA.
AC P57414;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ompa-like protein precursor.
GN OMPA OR B0332.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- FUNCTION: ACTS AS A PORIN WITH LOW PERMEABILITY THAT ALLOWS SLOW
CC PENETRATION OF SMALL SOLUTES (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (PROBABLY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AP001119; BAB3037.1; -
DR HSSP: P02934; I0JP.
DR InterPro: IPR001145; Bac_Ompa.
DR InterPro: IPR000498; Ompa_tmcm.
DR Pfam: PF00691; Ompa; 1.
DR Pfam: PF01389; Ompa_membrane; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_Ompa; 1.
DR PROSITE: PS01068; OMPA; FALSE_NEG.
KW Transmembrane; Porin; Signal; Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 349 OMPA-LIKE PROTEIN.
FT TRANSMEM 27 40 POTENTIAL.
FT TRANSMEM 63 75 POTENTIAL.
FT TRANSMEM 78 93 POTENTIAL.
FT TRANSMEM 106 116 POTENTIAL.
FT TRANSMEM 120 135 POTENTIAL.
FT TRANSMEM 154 165 POTENTIAL.
FT TRANSMEM 171 187 POTENTIAL.
FT TRANSMEM 193 204 POTENTIAL.
FT DOMAIN 207 218 HINGE LIKE.
FT DOMAIN 267 311 OMPA-LIKE.
FT DISULFID 321 333 BY SIMILARITY.
SQ SEQUENCE 349 AA; 39303 MW; CC14AB1BD590CF58 CRC64;
Query Match 15.7%; Score 135.5; DB 1; Length 349;
Best Local Similarity 31.2%; Pred. No. 4.9e-05;
Matches 34; Conservative 23; Mismatches 45; Indels 7; Gaps 3;

RESULT 27
MOTY_VIBAN

ID MOTY_VIBAN STANDARD: PRT; 293 AA.
AC Q9S3P9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sodium-type flagellar protein motY precursor (Polar flagellum motor
DE protein).
GN MOTY.
OS Vibrio anguillarum (Listonella anguillarum).
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Listonella.
OX NCBI_TaxID=55601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NB10 / Serotype O1;
RA Ormande P., Milton D.L.;
RT "The role of motility in adherence and invasion of a fish cell line by
RT Vibrio anguillarum";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS THE ROLE OF A STATOR IN THE SODIUM FLAGELLAR
CC MOTOR, STABILIZING THE FORCE-GENERATING UNIT THROUGH DIRECT
CC INTERACTION WITH THE CELL WALL. INVOLVED IN ADHERENCE AND INVASION
CC OF FISH CELL LINE.
CC -1- SUBCELLULAR LOCATION: Membrane-associated (potential).
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF176946; AAD51752.1; -
DR InterPro: IPR001145; Bac_Ompa.
DR InterPro: IPR001035; MotY.
DR Pfam: PF00691; Ompa; 1.
DR Pfam: PF01023; NAFLMOTY.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_Ompa; 1.
DR PROSITE: PS01068; OMPA; FALSE_NEG.
KW Flagella; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 293 SODIUM-TYPE FLAGELLAR PROTEIN MOTY.
FT DOMAIN 220 264 OMPA-LIKE.
SQ SEQUENCE 293 AA; 33725 MW; 5618435B49F40D3B CRC64;
Query Match 14.8%; Score 127.5; DB 1; Length 293;
Best Local Similarity 28.6%; Pred. No. 0.00021;
Matches 30; Conservative 27; Mismatches 47; Indels 1; Gaps 1;
Db 68 SLVYPPDSDEIKRQAAILDEQAQFLTNQRA-RVLVAGHTDEKRSREYNMSEGERAV 126
Db 183 TLHYDRDNNVQMKASQKRAQIADYIRNODIDLVLSTYDVSQSRVSDLSERRAE 242
Qy 127 AVENYLLGKINOASVEIISFGGEPIJAGTNEASQNRRLS 171
Db 243 VLDRYFKSLGPELDRIQVGYGKRRPIADNASIGDKRRRYVIS 287
RESULT 28
MOTY_VIBPA
ID MOTY_VIBPA STANDARD: PRT; 293 AA.
AC P46233;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Sodium-type flagellar protein motY precursor.
GN MOTY.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-BB22;
 RX MEDLINE-94292449; PubMed-8021208;
 RA McCarter L.L.;
 RL "Mot", a component of the sodium-type flagellar motor.";
 CC J. Bacteriol. 176:4219-4225(1994).
 CC -1- FUNCTION: PLAYS THE ROLE OF A STATOR IN THE SODIUM FLAGELLAR
 CC MOTOR, STABILIZING THE FORCE-GENERATING UNIT THROUGH DIRECT
 CC INTERACTION WITH THE CELL WALL.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (potential).
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U06949; AAA21574.1;
 CC InterPro: IPR001145; Bac.OmpA.
 CC InterPro: IPR001035; Mot.
 CC Pfam: PF00691; OmpA; 1.
 CC PRINTS: PR01023; NAF1GMOTY.
 CC PRODOM: PD000930; Bac.OmpA; 1.
 CC PROSITE: PS01068; OMPA; 1.
 CC KEGG: Flagella; Signal.
 CC FT SIGNAL 1 21 POTENTIAL.
 CC FT CHAIN 22 293 SODIUM-TYPE FLAGELLAR PROTEIN MOTY.
 CC FT DOMAIN 220 264 OMPA-LIKE.
 CC SQ SEQUENCE 293 AA; 33406 MW; FDFP3A11B5DCFA5 CRC64;
 CC -----
 CC Query Match 14.2%; Score 122.5; DB 1; Length 293;
 CC Best Local Similarity 27.6%; Pred. No. 0.0006; Indels 1; Gaps 1;
 CC Matches 29; Conservative 26; Mismatches 49;
 CC QY 68 SLVYFDSDEIKPQAAAILDEQAFLLTNGTA-RVLVAGHDERGSRNLSGERAV 126
 CC DB 183 TLHYEROGDLTKASKRRLAQLADYVRHNDIDLVTYTDSDGKESQSLERRAF 242
 CC QY 127 AVANNYLLGKINGASVEIISFGEERPAFGTNEANSQRRRLS 171
 CC DB 243 SLKTYFSLGLPEDRIQVGYGKRRIADNGTPIGCKNRRAVIS 287
 CC -----
 CC RESULT 29
 CC OMPA_SPRMA STANDARD; PRT; 359 AA.
 CC AC F04845;
 CC DT 13-AUG-1987 (Rel. 05, Created)
 CC DT 13-AUG-1987 (Rel. 05, last sequence update)
 CC DT 01-NOV-1995 (Rel. 32, last annotation update)
 CC DE Outer membrane protein A precursor.
 CC GN OMPA.
 CC OS Serratia marcescens.
 CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC OC Serratia.
 CC OX NCBI_TaxID=615;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC MEDLINE-85035845; PubMed-6092858;
 CC RA Braun G., Cole S.T.;
 CC RT "DNA sequence analysis of the Serratia marcescens ompA gene:
 CC RT implications for the organisation of an enterobacterial outer
 CC RT membrane protein.";
 CC RL Mol. Gen. Genet. 195:321-328(1984).
 CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (PROBABLE).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X00618; CAA25254.1;
 CC DR PIR; S07298; S07298.
 CC DR HSSP; P02934; 10UP.
 CC DR InterPro: IPR001145; Bac.OmpA.
 CC DR InterPro: IPR000498; OmpA_tmern.
 CC DR Pfam; PF00691; OmpA; 1.
 CC DR PRINTS: PR01389; OmpA_membrane; 1.
 CC DR PRODOM: PD000930; Bac.OmpA; 1.
 CC DR PROSITE: PS01068; OMPA; 1.
 CC DR Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
 CC KEGG: Porin.
 CC FT SIGNAL 1 21
 CC FT CHAIN 22 359 OUTER MEMBRANE PROTEIN A.
 CC FT TRANSMEM 27 40 POTENTIAL.
 CC FT TRANSMEM 62 74 POTENTIAL.
 CC FT TRANSMEM 77 92 POTENTIAL.
 CC FT TRANSMEM 104 114 POTENTIAL.
 CC FT TRANSMEM 118 133 POTENTIAL.
 CC FT TRANSMEM 154 165 POTENTIAL.
 CC FT TRANSMEM 171 187 POTENTIAL.
 CC FT TRANSMEM 193 204 POTENTIAL.
 CC FT DOMAIN 210 219 5 x 2 AA TANDEM REPEATS OF A-P.
 CC FT REPEAT 210 211 1.
 CC FT REPEAT 212 213 2.
 CC FT REPEAT 214 215 3.
 CC FT REPEAT 216 217 4.
 CC FT REPEAT 218 219 5.
 CC FT DOMAIN 268 312 OMPA-LIKE.
 CC FT DISULFD 322 336 BY SIMILARITY.
 CC SQ SEQUENCE 359 AA; 38426 MW; 13992A037C19758B CRC64;
 CC -----
 CC Query Match 14.2%; Score 122; DB 1; Length 359;
 CC Best Local Similarity 30.9%; Pred. No. 0.00084;
 CC Matches 34; Conservative 19; Mismatches 49; Indels 8; Gaps 2;
 CC QY 50 APLVNDYKALASKLPISLYFDSDEIKPQAAAILDEQAFLLT--NQTARVLVACH 107
 CC DB 218 APVVEYTKRFT-----LKSDFLEFNKSTLKABEQQALDOLYQLSMPDKSGSVVLGY 271
 CC QY 108 TDEGSRREYNMSIGERRAVVRNVLGKGINQASVEIISFGEERPAFGT 157
 CC DB 272 TDVAGSDQYNOKLSEORASVVDLYLSKGLPSDKISARGMGEDAVYGT 321
 CC -----
 CC RESULT 30
 CC TP50_TREPA STANDARD; PRT; 417 AA.
 CC AC P38369; Q56358;
 CC DT 01-OCT-1994 (Rel. 30, Created)
 CC DT 01-OCT-1994 (Rel. 30, last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, last annotation update)
 CC DE Outer membrane protein TP50 precursor (Antigen TP57).
 CC GN TP50 OR TP0292.
 CC OS Treponema pallidum.
 CC OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 CC OX NCBI_TaxID=160;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC STRAIN=Nichols;
 CC MEDLINE-94156449; PubMed-8112835;
 CC RA Hardham J.M., Stamm L.V.;

RT "Identification and characterization of the *Treponema pallidum* tps50 gene, an ompa homolog." Infect. Immun. 62:1015-1025(1994).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=Nichols.

RC MEDLINE=95349399; PubMed=7623668;

RA Cox D.L., Atkins D.R., Porcella S.F., Norgard M.V., Radolf J.D.;

RT "Treponema pallidum in gel microdroplets: a novel strategy for investigation of treponemal molecular architecture." Mol. Microbiol. 15:1151-1164(1995).

RL [3]

RN SEQUENCE FROM N.A.

RP STRAIN=Nichols.

RC MEDLINE=96332770; PubMed=9665876;

RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,

RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Skolzen E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khatak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,

RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

RA Venter J.C.;

RT "Complete genome sequence of *Treponema pallidum*, the syphilis spirochete." Science 281:375-388(1998).

RL [1]

CC -1- FUNCTION: COULD FUNCTION AS A PORIN.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane (Probable).

CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U02628; AAA0517.2; -

DR EMBL: L28427; AAA9311.1; -

DR EMBL: AE001209; AAC65272.1; -

DR TIGR: TP0292;

DR InterPro: IPR001145; Bac_Ompa.

DR Pfam: PF00691; Ompa; 1.

DR ProDom: PD000930; Bac_Ompa; 1.

DR PROSITE: PS01068; OMPA; 1.

KW Transmembrane; Porlin; Signal; Outer membrane; Antigen;

KW Complete proteome.

FT STGM 1 26 POTENTIAL.

FT CHAIN 27 417 OUTER MEMBRANE PROTEIN TPNSO.

FT DOMAIN 345 390 OMPA-LIKE.

FT CONFLICT 127 127 MISSING (IN REF. 2).

FT CONFLICT 350 350 A -> E (IN REF. 2).

SO SEQUENCE 417 AA; 47608 MW; 079F579CFDD7288C CRC64;

Query Match 13.6%; Score 117; DB 1; Length 417;

Best Local Similarity 27.0%; Pred. No. 0.0029;

Matches 44; Conservative 22; Mismatches 65; Indels 32; Gaps 4;

QY 39 TGYTGVIYGYVA-----PLYVDNFKVAKLSKLPSTL----- 69

DB 253 SGGTLYRQGTATKANPFRPPDPARTVYQETLKLD--HMPDAKRVTEEGVTISTEN 310

QY 70 YFDPDSDEIRFOAAAIIDEOAFLTNQTARVAVAGHTBERGSGRENMISLGRRAVAYR 129

DB 311 VQFDGASASLAPSEVEKRLTAIELLRAFPDRELIVSCHAARRGSVQOQRISEBRADYVA 370

QY 130 NYLIGKS--INQASVEIISFGEEPIAFGTNEANSQNRRAELS 171

DB 371 RYIQLGVADAHAHYTRCGAQOOSIAINDSEDRKRRRVEIT 413

RESULT 31

MOTB_BACSU

ID MOTB_BACSU STANDARD: PRT: 261 AA.

AC P28612;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Chemotaxis motB protein (Motility protein B).

GN MOTB.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_Taxid=1423;

RP [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=92325000; PubMed=1624413;

RA Mirel D.B., Lustre V.M., Chamberlin M.J.;

RT "An operon of *Bacillus subtilis* motility genes transcribed by the sigma D form of RNA polymerase." J. Bacteriol. 174:4197-4204(1992).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=168;

RC MEDLINE=96044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borries R., Boutsier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Gollighly E.J., Grandi G.,

RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hult M.F., Ilaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Kraerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestel D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

RA Priesgen E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivoita C., Roche A., Roche B., Rose M., Sadate Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scifone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,

RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpilstra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassartoli A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,

RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*." Nature 390:249-256(1997).

RL Nature 390:249-256(1997).

CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE CELL WALL.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein.

CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M77238; AAA22603.1; -

DR EMBL: Z99111; CAB13241.1; -

DR PIR: S27516; S27516.

DR PIR: B42882; B42882.

DR Subtilist; BG10689; motB.

DR InterPro: IPR001145; Bac_Ompa.

DR Pfam: PF00691; OmpA; 1.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Chemotaxis; Flagella; Transmembrane; Flagellar rotation;
 FT Complete proteome.
 FT DOMAIN 1 19 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 20 41 POTENTIAL.
 FT DOMAIN 261 261 EXTRACELLULAR (POTENTIAL).
 FT SEQUENCE 261 AA; 29482 MW; 382382A64C6D5 CRC64;
 SQ

Query Match 12.4%; Score 106.5; DB 1; Length 261;
 Best Local Similarity 20.6%; Pred. No. 0.014; 79; Indels 71; Gaps 7;
 Matches 47; Conservative 31; Mismatches 79; Indels 71; Gaps 7;

1 MMLHIOIAAAALSVLFMTGCAKSTQVWVAPNAPGYGVY-ITGVAD----- 51
 34 IYLVASSIDAARQML-----SKSFNEVF-----TGTGVLDSVTPPENESDGI 80
 FT
 FT NON_TER 243 243
 FT SEQUENCE 243 AA; 25972 MW; 687CC9C10DDEC64C CRC64;
 SQ

Query Match 12.1%; Score 104; DB 1; Length 243;
 Best Local Similarity 29.1%; Pred. No. 0.022;
 Matches 34; Conservative 16; Mismatches 55; Indels 12; Gaps 2;

66 LPSLVFDESDKIPQAAALIDEOQFLTT--NOTARVAVGHDERGSRREYNLSGER 123
 122 LKSDVLENSAKSSLIKPGQALDLYTQUSMDPKGVSVALGYIDPVKDAANKLSEA 181
 FT
 FT NON_TER 243 243
 FT SEQUENCE 243 AA; 25972 MW; 687CC9C10DDEC64C CRC64;
 SQ

124 RAVAVRNYLLGKGINQASV-EIISGEEPIAGCTNEAMSQNRRAEL 170
 201 RAVNFMGLLIENPKLDAKVFSAKGYGKPYASNKTAEGRSKNRREV 248
 DB

RESULT 32
 OMPA_SEROD STANDARD: PRT; 243 AA.
 ID OMPA_SEROD STANDARD: PRT; 243 AA.
 AC P24755;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Outer membrane protein A (outer membrane protein II) (Fragment).
 GN OMPA.
 OS Seratia odorifera.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Seratia.
 OX NCBI_TaxId=618;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-ATCC 3307;
 RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria."
 RT J. Gen. Microbiol. 137:1911-1921(1991).
 CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVER LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (PROBABLY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M63357; AAA26561.1; -
 CC HSSP: P02934; 1QJP.
 CC InterPro: IPR001145; Bac_OmpA.
 CC InterPro: IPR000498; OmpA_tmem.
 CC Pfam: PF00691; OmpA; 1.
 DR Pfam: PF01389; OmpA_membrane; 1.

DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; FALSE_NEG.
 KW Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.
 FT NON_TER 1 1
 FT TRANSMEM 12 8 POTENTIAL.
 FT TRANSMEM 12 27 POTENTIAL.
 FT TRANSMEM 48 59 POTENTIAL.
 FT TRANSMEM 65 81 POTENTIAL.
 FT TRANSMEM 87 98 POTENTIAL.
 FT DOMAIN 104 113 5 X 2 AA TANDEM REPEATS OF A-P.
 FT REPEAT 104 105 1.
 FT REPEAT 106 107 2.
 FT REPEAT 108 109 3.
 FT REPEAT 110 111 4.
 FT REPEAT 112 113 5.
 FT DISULFID 215 229 BY SIMILARITY.
 FT NON_TER 243 243
 FT SEQUENCE 243 AA; 25972 MW; 687CC9C10DDEC64C CRC64;
 SQ

Query Match 12.1%; Score 104; DB 1; Length 243;
 Best Local Similarity 29.1%; Pred. No. 0.022;
 Matches 34; Conservative 16; Mismatches 55; Indels 12; Gaps 2;

66 LPSLVFDESDKIPQAAALIDEOQFLTT--NOTARVAVGHDERGSRREYNLSGER 123
 122 LKSDVLENSAKSSLIKPGQALDLYTQUSMDPKGVSVALGYIDPVKDAANKLSEA 181
 FT
 FT NON_TER 243 243
 FT SEQUENCE 243 AA; 25972 MW; 687CC9C10DDEC64C CRC64;
 SQ

124 RAVAVRNYLLGKGINQASV-EIISGEEPIAGCTNEAMSQNRRAEL 170
 201 RAVNFMGLLIENPKLDAKVFSAKGYGKPYASNKTAEGRSKNRREV 248
 DB

RESULT 33
 YTXE_BACME STANDARD: PRT; 218 AA.
 ID YTXE_BACME STANDARD: PRT; 218 AA.
 AC P46827;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 24.6 kDa protein in CCPA 3 region (ORF2).
 GN YTXE.
 OS Bacillus megaterium.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxId=1404;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94259294; PubMed=8200532;
 RA Hueck C., Kraus A., Hillen W.;
 RT "Sequences of ccpA and two downstream Bacillus megaterium genes with
 RT homology to the molB operon from Bacillus subtilis."
 RT Gene 143:147-148(1994).
 CC -1- FUNCTION: MAY BE INVOLVED IN SOME TRANSPORT FUNCTION.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: L26052; AAA22297.1; -
 CC InterPro: IPR001145; Bac_OmpA.
 CC Pfam: PF00691; OmpA; 1.
 CC ProDom: PD000930; Bac_OmpA; 1.
 CC Hypothetical protein; Transport; Transmembrane.
 CC CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 1 14
 CC TRANSMEM 15 35
 CC TRANSMEM 36 218
 CC DOMAIN 218 AA; 24642 MW; 63622D730A4AE247 CRC64;
 FT
 FT SEQUENCE 218 AA; 24642 MW; 63622D730A4AE247 CRC64;
 SQ


```

CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U28219; AAB61254.1; -
CC DR EMBL: AE001244; AAC65689.1; -
CC DR TIGR: TP0724; -
CC DR InterPro: IPR001145; Bac_OmpA.
CC DR Pfam: PF00691; OmpA; 1.
CC DR ProDom: PD000930; Bac_OmpA; 1.
CC DR Chemotaxis; Flagella; Transmembrane; Inner membrane;
CC Flagellar rotation; Complete proteome.
CC FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 15 34 POTENTIAL.
CC FT DOMAIN 35 238 PERIPLASMIC (POTENTIAL).
CC SQ SEQUENCE 238 AA; 26050 MW; 0AB5FCDBE9852E1F CRC64;

Query Match
Best Local Similarity 29.4%; Score 95.5; DB 1; Length 238;
Matches 37; Conservative 20; Mismatches 58; Indels 11; Gaps 5;

QY 53 VDNDETVKALASKLPVLYFDFDSEIK-PQAAAILDQAOPLTTNQA--RVLVAGHTD 109
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 YVSD--RGLVYISLTSDFYFGSSDLNWEESREALLRVAQFLSHALAGRFRFEGHTD 165
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 -----ERGSREYNMSLGERRAVAVRNLLGKINGOASVEIISFGEERPIAGTNEAMSQ 164
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 SVEVPEDGSTD-NMELSTRRAVRVLYLTDFGAQNRSLAGYADTRAFKFSNESSEGRAY 224
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 165 NRRAPL 170
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 NRRVVD 230

RESULT 36
MOTB_HELPY STANDARD; PRT; 257 AA.
AC P56427;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chemotaxis motB protein (Molilly protein B).
GN MOTB OR HP0816.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clanton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodok A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Vertes U.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT
CC BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE
CC CELL WALL (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.

```

```

CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000593; AAD07865.1; -
CC DR TIGR: HP0816; -
CC DR InterPro: IPR001145; Bac_OmpA.
CC DR Pfam: PF00691; OmpA; 1.
CC DR ProDom: PD000930; Bac_OmpA; 1.
CC DR Chemotaxis; Flagella; Transmembrane; Inner membrane;
CC Flagellar rotation; Complete proteome.
CC FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 17 37 POTENTIAL.
CC FT DOMAIN 38 257 PERIPLASMIC (POTENTIAL).
CC SQ SEQUENCE 257 AA; 28849 MW; 58C1774B0547C032 CRC64;

Query Match
Best Local Similarity 20.9%; Score 87; DB 1; Length 257;
Matches 43; Conservative 27; Mismatches 88; Indels 48; Gaps 6;

QY 13 ALSVTFMTGCAKSTSQ-----VWAPNAPPGYGVYITGYA 50
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29 ALPIALYLAISAVNKSVKALTEFIKFNYPAPKPEAMQPVVYIPDSGKEEDOMSESSK 88
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 51 PLVNDDETVKALAS-----KLPDS-LVYDFPDSEIKPQAAAILDQAO 93
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 89 PASQNETETKATIKRKGEVSLQIDGSLIKLSNLFENASDAILNOMMLYIEIKAKI 148
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 L-TTNOTARVLVAGHTE-----RGSREYNMSLGERRAVAVRNLLGKINGOASVEIIS 146
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 IQKLPKRVHINVGFTDDPTLVTKREKSH--ELANRAVRVVKVLIQGVNPNQISFSS 206
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 FGEERPIAGTNEEAMSONRRAPLSY 172
   111 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 YGNTPIAPNDSLENRKNNRVEIFF 232

RESULT 37
YTXE_BACSU STANDARD; PRT; 242 AA.
AC P39064;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein YTXE.
GN YTXE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020526; PubMed=7934817;
RA Grundy F.J., Waters D.A., Takova T.Y., Henkin T.M.;
RA "Identification of genes involved in utilization of acetate and
RA acetoin in Bacillus subtilis."
RL Mol. Microbiol. 10:259-271(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Gallerton N., Sorokin A., Ehrlich S.D.;
RA "Sequencing and functional annotation of the Bacillus subtilis genes
RA in the 200 kb rmb-dnaB region."
RL Microbiology 143:3431-3441(1997).
RN [3]
RP SEQUENCE FROM N.A.

```

RC STRAIN-168;
 RA MEDLINE-98044033; PubMed-9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcher S.,
 RA Boriss R., Bourier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Ertlan K.D., Erington J., Fabret C., Ferrati E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holst S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priesen E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivalet C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scifone F.,
 RA Sekiguchi J., Sekowska A., Serr S.J., Serron P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Vandenbol M., Vannier F., Vassart A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,
 RA Viari A., Wambat R., Wedler E., Wedler H., Weltzeneger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein E.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN SOME TRANSPORT FUNCTION.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (potential).
 CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L17309; AAA68283.1;
 DR EMBL: AF008220; AAC00301.1;
 DR EMBL: 299119; CAB14950.1;
 DR PIR: S39642; S39642.
 DR Subtilisin; BG10366; YLXF.
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
 FT DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 24 44 POTENTIAL.
 FT DOMAIN 45 242 EXTRACELLULAR (POTENTIAL).
 FT SEQUENCE 242 AA; 27595 MW; 88A7DDC103CIDFE5 CRC64;

Query Match 10.0%; Score 86; DB 1; Length 242;
 Best Local Similarity 26.4%; Pred. NO. 0.94; Indels 6; Gaps 3;
 Matches 29; Conservative 13; Mismatches 62;

QY 66 LPSLVYFDSDDEIKPQAAALIDQOFLTNQTAFLVAGHTDERSGREY---NNSTG 121
 DB 124 LQEAFLFTGEAKVLLKNETLLHGLAVLQITPN-DIQVEGHDTSRNISTRTYSNNELS 182
 QY 122 ERRAAVAVNYLGR-GINQASVELISGEERPIAFGNEEAWSONRAEL 170
 DB 183 AARAGVLYFTSKELPSKRFIAVGYADTRKVPADNKTNEHMKENRRVEI 232

RESULT 38

G6P2_STRCO
 ID G6P2_STRCO STANDARD; PRT; 551 AA.
 AC 095523;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Glucose-6-phosphate isomerase 2 (EC 5.3.1.9) (GPI 2) (phosphoglucose
 DE isomerase 2) (Prt 2) (Phosphohexose isomerase 2) (Prt 2).
 GN PG12 OR SC01942 OR SC054.02C.
 OS Streptomyces coelicolor.
 CC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
 CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 CC NCBI_TaxID:1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RX MEDLINE-21996410; PubMed-1200953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrrell B.G., Parkhill J.,
 RT Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)".
 RL Nature 417:141-147(2002).
 CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-
 CC phosphate.
 CC -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AL035591; CAB38132.1;
 DR HSSP: Q9N1E2; IHOX.
 DR InterPro: IPR001672; G6P_Isomerase.
 DR Pfam: PF00342; PGI; 1.
 DR PRINTS: PR00662; G6PISOMERASE.
 DR PROSITE: PS00765; P_GLUCOSE_ISOMERASE_1; 1.
 DR PROSITE: PS00174; P_GLUCOSE_ISOMERASE_2; 1.
 KW Isomerase; Gluconeogenesis; Glycolysis; Complete proteome.
 FT ACT_SITE 390 390 BY SIMILARITY.
 FT ACT_SITE 514 514 BY SIMILARITY.
 FT SEQUENCE 551 AA; 60569 MW; 631BD4A76EB34643 CRC64;

Query Match 9.6%; Score 83; DB 1; Length 551;
 Best Local Similarity 25.0%; Pred. NO. 4.7;
 Matches 23; Conservative 14; Mismatches 53; Indels 2; Gaps 1;

QY 30 QVAVAPAPPGYGVYITGVAPLVNDDEFYKALASLPSLVYFDSDDEIKPQAAALIDE 89
 DB 69 QELAAATGVSGLDAMFRGERINTEDRAVLHTALRAPDAVLEGVENVVQVHAVLDK 128
 QY 90 QAOFLTNQTAFLVAGHTDERSGREYNNSTG 121
 DB 129 MAGFADRVKGE--WTGHTGRRIRNVYVNGIG 158

RESULT 39
 MOTB_HELPJ STANDARD; PRT; 257 AA.
 ID MOTB_HELPJ
 AC 092129;
 DT 16-OCT-2001 (Rel. 40, Created)

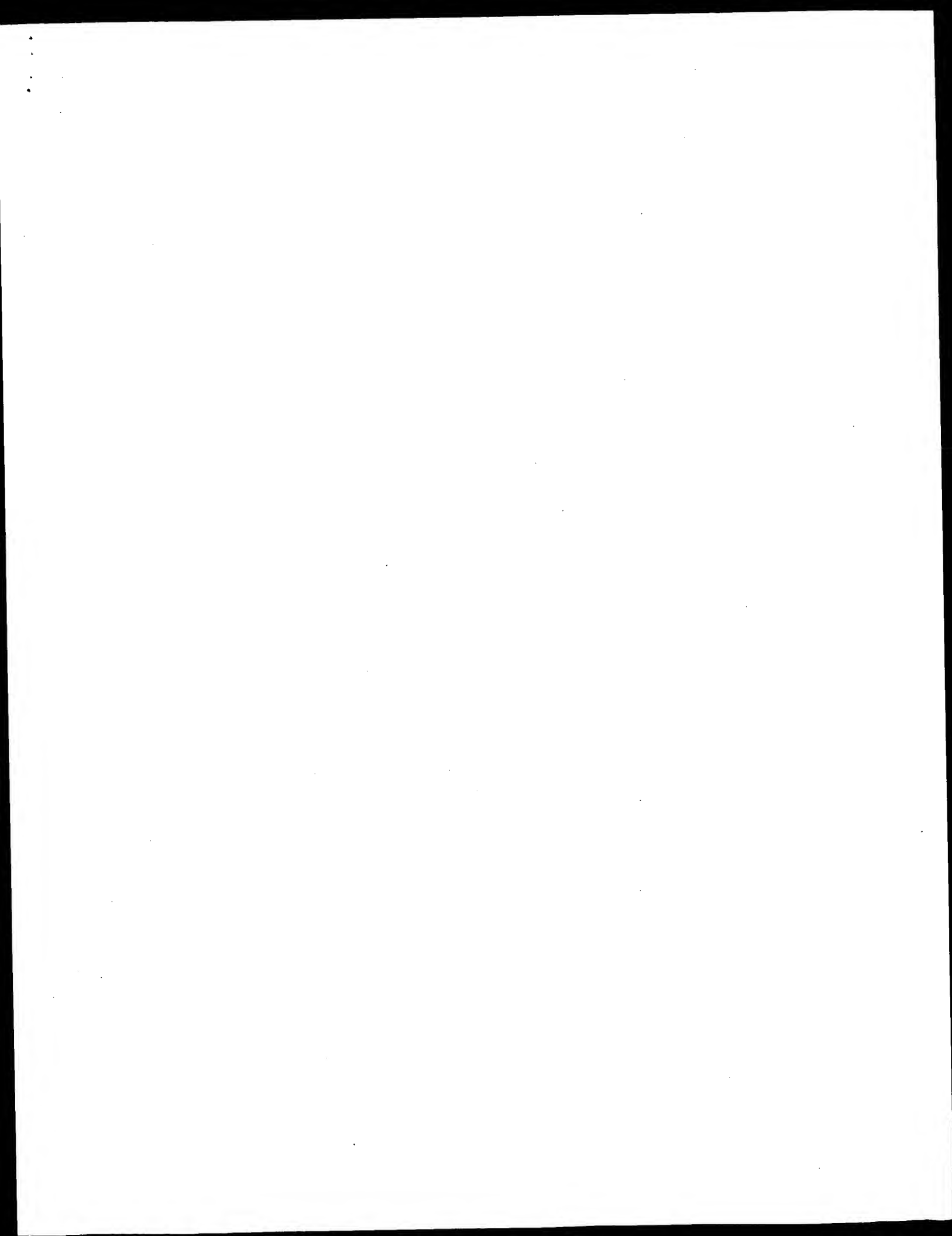
DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Chemotaxis motB protein (Motility protein B).
 GN MOTB OR JHP0752.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RX NCBI_TaxID=85963;
 RP [1]
 RX SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummiano P.J., Caruso A., Uria-Nikelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 387:176-180(1999).
 CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT
 CC BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE
 CC CELL WALL (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AEO01506; AAD06331.1; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA: 1.
 DR ProDom: PD000930; Bac_OmpA: 1.
 KW Chemotaxis; Flagella; Transmembrane; Inner membrane;
 KW Flagellar rotation; Complete proteome.
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 37 POTENTIAL.
 FT DOMAIN 38 257 PERIPLASMIC (POTENTIAL).
 FT SEQUENCE 257 AA; 28838 MW; 04F8E7BF833A9FB CRC64;
 SQ
 Query Match 9.5%; Score 82; DB 1; Length 257;
 Best Local Similarity 20.4%; Pred. No. 2.3;
 Matches 42; Conservative 25; Mismatches 91; Indels 48; Gaps 5;
 QY 13 ALSVLFMTGCAKSTSQ-----VMAAPAPGYGYITGYA 50
 DB 29 ALFTALYSAVSKSVKAEKTEFTIKFENYAPKPEAMOPVVIIPDGSKEEDOMASESK 88
 QY 51 PLVNDDETVALASKLPILVYFPDSDDEIKPQAAILDQAOFLITNQ-----TARV 102
 DB 89 PASQNETKATIKAKGSGVLEQIDGSAVLKLPSSILFENASDAINQOMLIYETIAT 148
 QY 103 L-----VAGHDE-----KGSREYNMSLGERRAVAVRNLGKINGQASVEITS 146
 DB 149 IQKLPRKHINVGFTDNTPLNKTREKSHY--ELANRAVRYVKVKVLIQGVDPNQISFSS 206
 QY 147 FGEERPIATGTNEEAMSONRAELSY 172
 DB 207 YGSTNPAPNDLSLENRRKNRYEIF 232
 RESULT 40
 COOS_RHORI STANDARD; PRT; 639 AA.
 AC P31896;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Carbon monoxide dehydrogenase (EC 1.2.99.2) (CODH).

GN COOS.
 OS Rhodospirillum rubrum.
 CC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 CC Rhodospirillum.
 RX NCBI_TaxID=1085;
 RP [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP STRAIN-URL;
 RX MEDLINE=92355502; PubMed=1644755;
 RA Kerby R.L., Hong S.S., Ensign S.A., Coppoc L.J., Ludden P.W.,
 RA Roberts G.P.;
 RT "Genetic and physiological characterization of the Rhodospirillum
 RT rubrum carbon monoxide dehydrogenase system.";
 RL J. Bacteriol. 174:5284-5294(1992).
 RL [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN-URL;
 RX MEDLINE=96198151; PubMed=862676;
 RA Fox J.D., Kerby R.L., Roberts G.P., Ludden P.W.;
 RT "Characterization of the CO-induced, CO-tolerant hydrogenase from
 RT Rhodospirillum rubrum and the gene encoding the large subunit of the
 RT enzyme.";
 RL J. Bacteriol. 178:1515-1524(1996).
 CC -1- FUNCTION: CODH MAY OXYDATE CARBON MONOXIDE COUPLED, VIA COOF, TO
 CC REDUCTION OF HYDROGEN CATION BY AN HYDROGENASE (POSSIBLY COOH).
 CC -1- CATALYTIC ACTIVITY: CO + H(2)O + acceptor = CO(2) + reduced
 CC acceptor.
 CC -1- COFACTOR: IRON, NICKEL AND ZINC.
 CC -1- INDUCTION: BY carbon monoxide; under anaerobic conditions.
 CC -1- MISCELLANEOUS: METHYL VIOLOGEN CAN ACT AS ACCEPTOR.
 CC -1- SIMILARITY: STRONG, TO C.THERMILACTICUM CARBON MONOXIDE
 CC DEHYDROGENASE BETA SUBUNIT AND SOME, TO M.SOEHNGENII CARBON
 CC MONOXIDE DEHYDROGENASE ALPHA SUBUNIT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U65510; AAC5123.1; -
 DR PIR: C42957; C42957.
 DR InterPro: IPR004137; Prismane.
 DR Pfam: PF03063; Prismane: 1.
 KW Oxidoreductase; Nickel; Iron; Zinc; Iron-sulfur.
 FT METAL 50 50 IRON-SULFUR (BY SIMILARITY).
 FT METAL 53 53 IRON-SULFUR (BY SIMILARITY).
 FT METAL 72 72 IRON-SULFUR (BY SIMILARITY).
 FT SEQUENCE 639 AA; 66854 MW; DCADDTC13D8D588 CRC64;
 SQ
 Query Match 9.4%; Score 81; DB 1; Length 639;
 Best Local Similarity 22.6%; Pred. No. 8.6;
 Matches 42; Conservative 26; Mismatches 58; Indels 60; Gaps 8;
 QY 4 HIQIAAAALSVLEMTGCAKSTSQV--WAPAPAPGYGYITGY-----AP 51
 DB 462 YVDLAKSLAKRNVLVLATGCAAGAFAGAGMTSEATQYAGELKGVLSAIGTAAGIGCP 521
 QY 52 L-----VDNETYKALASKLPILVYFPDSDDEIKPQAAILDQO----- 91
 DB 522 LPLVHMGSQVDSNRVAV-ALATVALANKIGVDSLPLVASPECSKALALIGSWAVTIG 580
 QY 92 -----OFLT--TNOTARVYLVAQH-----TDERGSEYNNMSLGERRAVAVRNY 131
 DB 581 LPTHWGSPVPIYISQIYIKVLTETAKDLVGSYFLVDIDPK-----SADKLYAAIQER 633
 QY 132 LLGKI 137
 DB 634 RAGIGL 639

Mon Jul 7 08:55:32 2003

Search completed: July 6, 2003, 14:05:18
Job time : 39 secs

us-09-674-779b-2.rsp



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:04:35 / Search time 58 Seconds

(without alignments)
611.036 Million cell updates/sec

Title: US-09-674-779B-2

Sequence: 1 MMLHIQTAAALSVTFM.....IAFGTNEAMSONRRALISY 172

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Database:

Post-processing: Minimum Match 10%

Maximum Match 100%

Listing first 100 summaries

SPREMBL_21:*

1: sp.archaea:*

2: sp.bacteria:*

3: sp.fungi:*

4: sp.human:*

5: sp.invertebrate:*

6: sp.mammal:*

7: sp.mhc:*

8: sp.organelle:*

9: sp.phage:*

10: sp.plant:*

11: sp.todent:*

12: sp.virus:*

13: sp.vertibrate:*

14: sp.unclassified:*

15: sp.virus:*

16: sp.bacterioph:*

17: sp.archaeoph:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	307	35.7	168	2	051489	
2	306	35.5	168	16	091424	051489 pseudomonas
3	289	33.6	174	16	08XfH6	08XfH6 pseudomonas
4	287.5	33.4	168	2	093465	08XfH6 salmonella
5	286.5	33.3	172	16	09A1F4	093465 erwina chr
6	286.5	33.3	188	16	09A3H5	09A1F4 raistonia s
7	283	32.9	168	16	08ZG20	09A3H5 caulobacter
8	269.5	31.3	168	16	09B8F5	08ZG20 yersinia pe
9	265.5	30.8	186	16	09PC85	09B8F5 rhizobium l
10	262.5	30.5	176	16	09Z6C3	09PC85 xyella fas
11	262	30.4	177	16	08U9L5	09Z6C3 rhizobium m
12	259	30.1	172	16	09KR12	08U9L5 agrobacteri
13	258	30.0	130	2	09L308	09KR12 vibrio chol
14	256	29.7	155	16	09ZCH2	09L308 pasteurella
15	254	29.5	168	16	044662	09ZCH2 rickettsia
16	253.5	29.4	152	2	044157	044662 brucella ab
						044157 actinobacil

90 160 18.6 220 16 08XFE2
 91 160 18.6 326 2 006895
 92 160 18.6 340 2 09X4S2
 93 158.5 18.4 344 2 09X4L6
 94 158.5 18.4 344 2 08RML8
 95 158.5 18.4 344 2 08RJD0
 96 156.5 18.2 207 16 09KLT2
 97 156.5 18.2 407 2 P96774
 98 156 18.1 341 2 08R558
 99 155.5 18.1 341 2 08R558
 100 154.5 17.9 163 2 052584
 08xfe2 salmonella
 006895 campylobact
 09x4s2 pseudomonas
 09x4l6 pseudomonas
 08rml8 haemophilus
 08rjd0 haemophilus
 09klt2 vibrio chol
 0914t3 pseudomonas
 P96774 haemophilus
 08r558 haemophilus
 052584 pseudomonas

ALIGNMENTS

RESULT 1
 ID 051489 PRELIMINARY; PRT; 168 AA.
 AC 051489:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE Outer membrane protein.
 GN OPR1.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas
 NX NCBI_TaxID=287;
 RN
 RP SEQUENCE FROM N.A.
 RX LIM A., De Vos D., Brauns M., Mossialos D., Gaballa A., Hamers R.,
 RA Cornelis P.;
 RT "Molecular and immunological characterization of OprL, the 18 kDa
 RT outer-membrane peptidoglycan-associated lipoprotein (Pal) of
 RT Pseudomonas aeruginosa";
 RL Microbiology 143:1703-1716(1997).
 DR EMBL; Z50191; CAA90573.1; -;
 DR InterPro; IPR001145; Bac_OmpA.
 DR Pfam; PF00691; OmpA; 1.
 DR PRINTS; PR01021; OMPADOMAIN.
 DR ProDom; PD000930; Bac_OmpA; 1.
 DR PROSITE; PS01068; OMPA; 1.
 SQ SEQUENCE 168 AA; 17898 MW; 74831279BBA1186 CRC64;
 Query Match 35.7%; Score 307; DB 2; Length 168;
 Best Local Similarity 40.9%; Pred. No. 7.4e-21;
 Matches 70; Conservative 39; Mismatches 44; Indels 18; Gaps 6;
 QY 6 QIAAAALSVLTFTMGCAK-----STQVWVAPNAPGTGVTGYVAVPLVNDDET 59
 DB 8 KFAALALAMAV---AVGSSKSGDASGEGANGVDPA--GY-GANSAGVDSLSDEAL 61
 QY 60 KALASKLPILVYFEDSDDEIKPQAAALDEQAQFLTTNOTARVLVAGHTDEGSEYKMS 119
 DB 62 RAITT-----FYFYDSSDLKPEAKRALDVHAKDL-KSGGORYLVLEGHTDGTREYNMA 115
 QY 120 LGERRAVAVRNYLIGKGINQASVEIISFGERRIAGTDEASQNRRAEL 170
 DB 116 LGERRAKAVORYLVLOGVSPALVELVSGKERPAVATGHDQSMQNRRAEL 166
 RESULT 2
 ID 091424 PRELIMINARY; PRT; 168 AA.
 AC 091424:
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Outer membrane protein OprL precursor.
 GN OPR1 OR PA0973.
 OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 NX NCBI_TaxID=287;
 RN
 RP SEQUENCE FROM N.A.
 RX STRAIN-ATCC 15692 / PA01;
 RC MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Labig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004530; AAG04362.1; -;
 DR InterPro; IPR001145; Bac_OmpA.
 DR Pfam; PF00691; OmpA; 1.
 DR PRINTS; PR01021; OMPADOMAIN.
 DR ProDom; PD000930; Bac_OmpA; 1.
 DR PROSITE; PS01068; OMPA; 1.
 KW Complete proteome.
 SQ SEQUENCE 168 AA; 17925 MW; 03E36279BBA1183 CRC64;
 Query Match 35.5%; Score 306; DB 16; Length 168;
 Best Local Similarity 40.9%; Pred. No. 9.2e-21;
 Matches 70; Conservative 38; Mismatches 45; Indels 18; Gaps 6;

QY 6 QIAAAALSVLTFTMGCAK-----STQVWVAPNAPGTGVTGYVAVPLVNDDET 59
 DB 8 KFAALALAMAV---AVGSSKSGDASGEGANGVDPA--GY-GANSAGVDSLSDEAL 61
 QY 60 KALASKLPILVYFEDSDDEIKPQAAALDEQAQFLTTNOTARVLVAGHTDEGSEYKMS 119
 DB 62 RAITT-----FYFYDSSDLKPEAKRALDVHAKDL-KSGGORYLVLEGHTDGTREYNMA 115
 QY 120 LGERRAVAVRNYLIGKGINQASVEIISFGERRIAGTDEASQNRRAEL 170
 DB 116 LGERRAKAVORYLVLOGVSPALVELVSGKERPAVATGHDQSMQNRRAEL 166
 RESULT 3
 ID 08XFE6 PRELIMINARY; PRT; 174 AA.
 AC 08XFE6:
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Tol protein required for outer membrane integrity, uptake of group A
 DE colicins, and translocation of phage DNA to cytoplasm
 DE (peptidoglycan-associated lipoprotein).
 GN PAL OR STY0749 OR STY0795.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella
 NX NCBI_TaxID=602, 601;
 RN
 RP SEQUENCE FROM N.A.
 RX SPECIES-S. typhimurium; STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RC MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2";
 RL Nature 413:852-856(2001).
 RN
 RP SEQUENCE FROM N.A.
 RX SPECIES-S. typhi; STRAIN-CT18;

MEDLINE-21534947; PubMed-11677608;
 RA ParKhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connetton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogan A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18.",
 RL Nature 413:848-852(2001).
 DR EMBL; AE008730; AAL19693.1;
 DR EMBL; AL627268; CAD05211.1;
 DR InterPro; IPR001145; Bac_Ompa.
 DR Pfam; PF00691; Ompa; 1.
 DR PRINTS; PRO1021; OMPADOMAIN.
 DR ProDom; PD000930; Bac_Ompa; 1.
 DR PROSITE; PS01068; OMPA; 1.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 174 AA; 1865 MW; A515624F96EB5A5 CRC64;

Query Match 33.6%; Score 289; DB 16; Length 174;
 Best Local Similarity 36.8%; Pred. No. 3.7e-19;
 Matches 67; Conservative 33; Mismatches 64; Indels 18; Gaps 3;

QY 1 MMLHIQTAAALSVLTFTGTCANKSTQVWVAPNAPGYTGVIYGVAPLVD----- 54
 DB 1 MMLHIQTAAALSVLTFTGTCANKSTQVWVAPNAPGYTGVIYGVAPLVD----- 54
 QY 55 --NDEYKALASLP--SLVYDFPDSDEIKFOAAALDEQAFLTTNQTARVLVAGHTDE 110
 DB 53 MSSEBOARLOMOQLQONNTVYFDLKYDIRSDFAMLDHAFLRNSPKYKTVGHADDE 112
 QY 111 RGSREYNMSLGERRAVAVNYLKGKINGQASVEIISFGEERPIAFGTNEEASQNRRAEL 170
 DB 113 RGPPEYNISLGERRAVAVNYLQKGVASADQISTYSGKEKPAVLGHDEAAVAKNRRAVL 172
 QY 171 SY 172
 DB 173 VY 174

RESULT 4
 0934G5 PRELIMINARY; PRT; 168 AA.
 AC 0934G5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL.
 OS *Erwinia chrysanthemi*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Peptobacterium.
 OX NCBI_TaxID=556;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3937;
 RA Ray M.C., Vianney A., Cotte-pattat N., Lazzaroni J.;
 RT "Characterization of the *Erwinia chrysanthemi* tol-pal genes."
 RT Submitted (Oct-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AJ297885; CAC82710.1;
 DR InterPro; IPR001145; Bac_Ompa.
 DR Pfam; PF00691; Ompa; 1.
 DR ProDom; PD000930; Bac_Ompa; 1.
 DR PROSITE; PS01068; OMPA; UNKNOWN_1.
 DR Lipoprotein; Signal.
 KW SIGNAL
 FT CHAIN 22 168 POTENTIAL.
 SQ SEQUENCE 168 AA; 18433 MW; 1E315BCF52852D55 CRC64;

Query Match 33.4%; Score 287.5; DB 2; Length 168;

Best Local Similarity 37.4%; Pred. No. 4.9e-19;
 Matches 65; Conservative 31; Mismatches 61; Indels 17; Gaps 3;
 QY 1 MMLHIQTAAALSVLTFTGTCANKSTQVWVAPNAPGYTGVIYGVAPLVDDEYK 60
 DB 10 LMLDLPLVLAAGSS-----NKRANNDOSSLNGAGMEN-----GNNSSAEQAR 54
 QY 61 ALASKLP--SLVYDFPDSDEIKFOAAALDEQAFLTTNQTARVLVAGHTDGRSREYN 118
 DB 55 LQWELQNNIVYFDLKYDIRDFOMLDAHAFLRNSPKYKTVGHADDEGTPPEYNI 114
 QY 119 SLGRRRAVAVNYLKGKINGQASVEIISFGEERPIAFGTNEEASQNRRAELSY 172
 DB 115 ALGERRAVAVNYLQKGVASADQISTYSGKEKPAVLGHDEAAVAKNRRAVLVY 168

RESULT 5
 08Y1F4 PRELIMINARY; PRT; 172 AA.
 ID 08Y1F4;
 AC 08Y1F4;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE Probable peptidoglycan-associated lipoprotein precursor.
 GN PAL OR RSC0736 OR RS05117.
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 CC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM1000;
 RX MEDLINE-21681879; PubMed-11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Atilat M., Billault A., Brotlier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choise N., Claudet-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T.,
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*,"
 RL Nature 415:497-502(2002).
 DR EMBL; AL646060; CAD14266.1;
 DR InterPro; IPR000104; Antifreeze_1.
 DR ProDom; PD000930; Bac_Ompa; 1.
 DR PRINTS; PRO0308; ANTIFREEZE1.
 DR ProDom; PD000930; Bac_Ompa; 1.
 DR PROSITE; PS01068; OMPA; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 172 AA; 18656 MW; 5CFE28E6AE7702D9 CRC64;

Query Match 33.3%; Score 286.5; DB 16; Length 172;
 Best Local Similarity 38.2%; Pred. No. 6.3e-19;
 Matches 65; Conservative 36; Mismatches 62; Indels 7; Gaps 3;

QY 5 IDIAAALSVLTFTGTCANKSTQVWVAPNAPGYTGVIYGVAPLVDNDEYKAL 62
 DB 8 IKLAIALDLADGACSSSVKLDITSK-----NATGAAAGADTRNTVYVDSHDELTPDN 62
 QY 63 ASKLP--SLVYDFPDSDEIKFOAAALDEQAFLTTNQTARVLVAGHTDGRSREYN 122
 DB 63 SPLAKRSYVDFDSTYKPEYOGILLTHARILQSHNRKYLIGQNTDRETSEYNALQO 122
 QY 123 RRAVAVNYLKGKINGQASVEIISFGEERPIAFGTNEEASQNRRAELSY 172
 DB 123 KRAEVRRAISLGVDPDSQMSVSLGKEKPAQASGHDESWAQRSDIY 172

RESULT 6
 09A3H5 PRELIMINARY; PRT; 188 AA.
 AC 09A3H5;
 ID 09A3H5;
 SQ SEQUENCE 188 AA; 18433 MW; 1E315BCF52852D55 CRC64;

DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE OmpA family protein.
 GN CC3229.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 NCBI_TaxID=155892;
 RM
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 MEDLINE=21173698; PubMed=11259647;
 RX Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry R.,
 RA Ullrich T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RL "Complete genome sequence of Caulobacter crescentus."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL: AE005987; AAC25191.1;
 DR TIGR: CC3229;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Complete proteome.
 SQ SEQUENCE 188 AA; 20425 MW; DF5703C15AEFB3B CRC64;

Query Match 33.3%; Score 286.5; DB 16; Length 188;
 Best Local Similarity 38.2%; Pred. No. 7.1e-19;
 Matches 65; Conservative 34; Mismatches 50; Indels 21; Gaps 4;

QY 7 IAAAAALSVLFMTGCANKSTQVAVAPNA-----PTGYT---GVITGVAPLVNDND 56
 DB 19 VGLAAASLA-----ACASRPKQPVTPPPAQPTQPTPTPPAPPVTSGLTP----- 67
 QY 57 ETVKALASKLPFLVYFDPSDEIRKQAAIIDEQAQFLTTNQTARVLVAGHTDGRSREY 116
 DB 68 GSVQDFVNVIGDVFDEYVIRADQAPVLAGQAKWMLRRSSVIRIEGNDERTREY 127
 QY 117 NNSIGERRAVAVRNYLLGKINQASVEIISFGEERPIAFGTNEAMSONR 166
 DB 128 NIALGARRNAVRFLLIAGVSSARLETISFKEKRPIDGSGSEAMAKNR 177

RESULT 7

08ZGZ0 PRELIMINARY; PRT; 168 AA.

AC 08ZGZ0;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Peptidoglycan-associated lipoprotein Pal.
 GN PAL OR YP01125.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 NCBI_TaxID=632;
 RM
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
 MEDLINE=21470413; PubMed=11586360;
 RX Parkhill J., Wren B.W., Thomson N.R., Tildall R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Farraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford B.G.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
 "Genome sequence of Yersinia pestis, the causative agent of plague."

RL Nature 413:523-527(2001).
 DR EMBL: AF414146; CAC89968.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; 1.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 168 AA; 18084 MW; 3AAA851E765F1994 CRC64;

Query Match 32.9%; Score 283; DB 16; Length 168;
 Best Local Similarity 37.9%; Pred. No. 1.3e-18;
 Matches 66; Conservative 32; Mismatches 68; Indels 8; Gaps 2;

QY 1 MMLHIOIAAAALSVLFMTGCANKSTQVAVAPNAFTGYTGVITGVAPLVNDDET 60
 DB 1 MQLNKVLCALMLALPYLVAACSSSKNSANNDSGAGTG-----TENGSLTSEEQAR 54
 QY 61 ALASKLP--SLVYFDESDERKQAAIIDEQAQFLTTNQTARVLVAGHTDGRSREY 118
 DB 55 LQMQLQKNNIYFFGDKYDIDSDFAQMDAAFAELRNSPSKVVVEGHADRGPEYNI 114
 QY 119 SLGERAVAVRNYLLGKINQASVEIISFGEERPIAFGTNEAMSONRRAELSY 172
 DB 115 ALGERASAVKMYLGKGVASDQISTVIGKEKPAVLGHDEAFAKRRRAVLVY 168

RESULT 8

098F85 PRELIMINARY; PRT; 168 AA.

AC 098F85;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Hypothetical protein ml13887.
 GN ML3887.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 NCBI_TaxID=381;
 RM
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=MAFF5303099;
 MEDLINE=21082930; PubMed=11214968;
 RX Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003003; BAB50682.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 168 AA; 17765 MW; 03E260D3BA1ECB3 CRC64;

Query Match 31.3%; Score 269.5; DB 16; Length 168;
 Best Local Similarity 40.3%; Pred. No. 2.3e-17;
 Matches 64; Conservative 25; Mismatches 53; Indels 17; Gaps 5;

QY 13 ALSVITFTGCANKSTQVAVAPN--APYGTGYIYIGVA--PLVDNDETVALASKLP 69
 DB 15 ALVAMLAATAGCAKSKT-----PNNADVLGNG--AAATPGSAQDFTV-----NIGDR 60
 QY 70 VYEDDSDEIRKQAAIIDEQAQFLTTNQTARVLVAGHTDGRSREYNSLGERRAVAVR 129
 DB 61 IFEDDSSIRDAQTTLARQAQWLNQKQYAIIVEGHADRGREYVALGARRAAAR 120
 QY 130 NYLLGKINQASVEIISFGEERPIAFGTNEAMSONRRA 168

Dh 121 DFLVSKGVASSRLKTSYGRKREPVAVCDISGMSQNRRA 159

RESULT 9

09PC85 PRELIMINARY; PRT; 186 AA.

AC 09PC85;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Outer membrane protein p6 precursor.
GN XFI896.

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;

OC Xylella.

NCBI_TaxID=2371;

RP SEQUENCE FROM N.A.

RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Aruda P., Abreu F.A., Acencio M.,
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
RA Colauto L.L., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Pelxoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
RA de Souza A.P., Terenzi M.L.Z., Siqueira W.J., de Souza A.A.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.",
RL Nature 406:151-159(2000).
DR EMBL: AE004009; AAF84702.1; -
DR InterPro: IPR001145; Bac_Ompa.
DR Pfam: PF00691; Ompa.1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_Ompa.1.
DR Complete proteome.
SQ SEQUENCE 186 AA; 20431 MW; C8DD2F6233DB0C92 CRC64;

Query Match 30.8%; Score 265.5; DB 16; Length 186;
Best Local Similarity 35.9%; Pred. No. 6.3e-17;
Matches 60; Conservative 30; Mismatches 54; Indels 23; Gaps 5;

OY 20 MTGANKSTSOVMV-----AP--NAPGYGYVITGYAPLVYNDDEVYKALASK 65
DB 26 LVASCKYKEQPHVPVKTMAPTVSTPAPTTAPDSSG-LYT--AAOLDIDACLR----- 77
OY 66 LPSLVYDFDSDEIKPQAAALIDBOAFLTTNOTARLVAGHTDERSREYNSLGERRA 125
DB 78 -QRYVYDFDKDVKKEQYVLGCHAKYLRNRPASHTLTGNDERSREYNALGRRG 136
OY 126 VAVANNYLKGINASVEIISFGERRIACGTNEANSQNRRELST 172
DB 137 NSVLISQANGASSGQNLNVSVSYGERPVCTESCSWNRNREIIV 183

RESULT 10

0926C3 PRELIMINARY; PRT; 176 AA.

AC 0926C3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Probable peptidoglycan-associated lipoprotein precursor.
GN PAL OR R02738 OR SMC02942.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
NCBI_TaxID=382;

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothé G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetalle D., Puehler A., Punelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.",
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL501791; GAC47317.1; -
DR InterPro: IPR001145; Bac_Ompa.
DR Pfam: PF00691; Ompa.1.
DR ProDom: PD000930; Bac_Ompa.1.
DR Complete proteome.

SW SEQUENCE 176 AA; 18735 MW; A6F162CB35042268 CRC64;
Query Match 30.5%; Score 262.5; DB 16; Length 176;
Best Local Similarity 38.5%; Pred. No. 1.1e-16;
Matches 60; Conservative 24; Mismatches 61; Indels 11; Gaps 3;

OY 13 ALSVTFMPCANKSTSOVMVAPNAPGYGYVITGYAPLVYNDDEVYKALASKPLVYF 72
DB 23 ALVWTLALACCAKSN-----LPNDAG-LGLAGAAATPGSQDFTV-----NVGDRIF 71
OY 73 DFDSEIKPQAAALIDBOAFLTTNOTARLVAGHTDERSREYNSLGERRAVAVRNTL 132
DB 72 DFDSTSRADAQATLDRQAQWLAKYPNYGTTIEGHADENGTREYNALATARRAAATRDYL 131
OY 133 LKGINASVEIISFGERRIACGTNEANSQNRRA 168
DB 132 VSRGVDPGNRRRTTSYGRKREPVAVCDISGMSQNRRA 167

RESULT 11

0809L5 PRELIMINARY; PRT; 177 AA.

AC 0809L5;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Omp16 protein.

GN P16A OR ATU0713 OR AGR_L_2246.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

NCBI_TaxID=176299;

RP SEQUENCE FROM N.A.

RX MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F.Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rose G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Ioo H., Tao Y., Biddle P., Jung K., Krespan W., Perry M.,

GN HP1125.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 NCBI_TaxID=210;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RA MEDLINE=9739467; PubMed=9252185;
 RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Olerbeck T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori.";
 RL Nature 388:539-547(1997).
 DR EMBL: AE000619; AAD08169.1; -
 DR TIGR: HP1125; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR Prodom: PD000930; Bac_OmpA.1.
 KW Hypothetical protein; Lipoprotein; Complete proteome.
 SQ SEQUENCE 179 AA; 19978 MW; 5B5521E284E27BAC CRC64;
 Query Match 25.0%; Score 215.5; DB 16; Length 179;
 Best Local Similarity 32.2%; Pred. No. 2.8e-12;
 Matches 56; Conservative 34; Mismatches 63; Indels 21; Gaps 5;
 QY 15 SVLTFF-----TGCANKSTFSQVWVAP-NAPTYGVITYT-----GVALYDND 56
 DB 5 SVFSFLVAFLLVYGCGHKMNKTAVAGDVSAKTQVATVTEPAPEKEEPKQEPAPVVEEK 64
 QY 57 ETVALASKLPSLVYDFDPSDEIKPQAAALIDBOAFLTTNGARVAVAGHTDERSREY 116
 DB 65 PAYE--SGTIIASIFDFDKYEIKESDQETLDEIVOKAKENH-MQVLLGNTDDEGSSSEY 121
 QY 117 NMSIGERRAAVAVNYLLGKINGASVEIISFGEERPIAGFTNEANSORRAEL 170
 DB 122 NOALGVKRTLSVKNALVIGVEKDMIKTISFGETKPKCAOKTRECYNRRVDV 175
 RESULT 21
 Q92K87 PRELIMINARY; PRT; 179 AA.
 ID Q92K87
 AC Q92K87;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Putative outer membrane protein.
 GN UHP1054.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 NCBI_TaxID=85963;
 [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=9120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jlang O., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 DR EMBL: AE001533; AAD06633.1; -
 DR InterPro: IPR001145; Bac_OmpA.

DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR Prodom: PD000930; Bac_OmpA.1.
 KW Complete proteome.
 SQ SEQUENCE 179 AA; 20030 MW; F62BB619186458BD CRC64;
 Query Match 25.0%; Score 215.5; DB 16; Length 179;
 Best Local Similarity 32.8%; Pred. No. 2.8e-12;
 Matches 57; Conservative 30; Mismatches 66; Indels 21; Gaps 5;
 QY 15 SVLTFF-----TGCANK-----STSQVWVAP-NAPTYGVITYTGAAPLYDND 56
 DB 5 SVFSFLVAFLLVYGCGHKMNKTAVAGDVSAKTQVATVTEPAPEKEEPKQEPAPVVEEK 64
 QY 57 ETVALASKLPSLVYDFDPSDEIKPQAAALIDBOAFLTTNGARVAVAGHTDERSREY 116
 DB 65 PAYE--SGTIIASIFDFDKYEIKESDQETLDEIVOKAKENH-MQVLLGNTDDEGSSSEY 121
 QY 117 NMSIGERRAAVAVNYLLGKINGASVEIISFGEERPIAGFTNEANSORRAEL 170
 DB 122 NOALGVKRTLSVKNALVIGVEKDMIKTISFGETKPKCAOKTRECYNRRVDV 175
 RESULT 22
 Q92AW7 PRELIMINARY; PRT; 179 AA.
 ID Q92AW7
 AC Q92AW7;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Omp22.
 GN Omp22.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 NCBI_TaxID=210;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KCTC0217BP;
 RA Seo W.Y., Kim J.S., Jang J.H., Yu G.J., Yum J.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U75869; AAD09577.1; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR Prodom: PD000930; Bac_OmpA.1.
 SQ SEQUENCE 179 AA; 20024 MW; 025ADC298F50C5AC CRC64;
 Query Match 24.2%; Score 208.5; DB 2; Length 179;
 Best Local Similarity 30.8%; Pred. No. 1.2e-11;
 Matches 52; Conservative 33; Mismatches 69; Indels 15; Gaps 4;
 QY 14 LSVLFFPMGCAKSTFSQVWVAP-NAPTYGVITYTGV-----APLYDNDIFYKA 61
 DB 10 LVAFLEFLVAGCGHKMNKTAVAGDVSAKTQVATVTEPAPEKEEPKQEPAPVVEEK PAYE- 68
 QY 62 LASKLPSLVYDFDPSDEIKPQAAALIDBOAFLTTNGARVAVAGHTDERSREYNNSTIG 121
 DB 69 -SGTIIASIFDFDKYEIKESDQETLDEIVOKAKENH-MQVLLGNTDDEGSSSEYNOALG 126
 QY 122 ERRAAVAVNYLLGKINGASVEIISFGEERPIAGFTNEANSORRAEL 170
 DB 127 VKRTLSVKNALVIGVEKDMIKTISFGETKPKCAOKTRECYNRRVDV 175
 RESULT 23
 Q8XWU0 PRELIMINARY; PRT; 209 AA.
 ID Q8XWU0
 AC Q8XWU0;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Probable transmembrane protein.

OY 104 VAGTDERGSRREYNSIGERRAVAVRNYLLGKGINQASVEIISPEERPIAFGTNEBAMS 163
 DB 95 VEGNCDEWGTDEYNOALGLKRAKAVKEALLAKGVNADRIAVKSGEINPCTEKTRACDA 154
 OY 164 QNRRAE 169
 DB 155 QNRRAE 160

RESULT 26

0927C5

ID 0927C5 PRELIMINARY; PRT; 192 AA.

AC 0927C5;

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

GN Peptidoglycan-associated lipoprotein.

OC Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=83558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;

RX MEDLINE=99206606; PubMed=10192388;

RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";

RL Nat. Genet. 21:385-389(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

RA White O., Hickey E.K., Peterson J., Ulfertback T., Berry K., Bass S.,

RA Linhar K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,

RT "Genome sequences of Chlamydia trachomatis MOP and Chlamydia

RL Pneumoniae AR39";

RN Nucleic Acids Res. 28:1397-1406(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=J138;

RX MEDLINE=20330349; PubMed=10871362;

RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,

RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,

RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

RL from Japan and CWL029 from USA";

RN Nucleic Acids Res. 28:2311-2314(2000).

DR EMBL: AE001659; AAD18919.1; -

DR EMBL: AE002265; AAF38862.1; -

DR EMBL: AF002547; BAA98989.1; -

DR TIGR: CP1091; -

DR InterPro: IPR001145; Bac_OmpA.

DR Pfam: PF00691; OmpA; 1.

DR ProDom: PD000930; Bac_OmpA; 1.

KW Lipoprotein; Complete proteome.

SQ SEQUENCE 192 AA; 21925 MW; A55094A12DD50D50 CRC64;

Query Match 23.8%; Score 204.5; DB 16; Length 192;
 Best Local Similarity 38.7%; Pred. No. 3.2e-11;
 Matches 46; Conservative 24; Mismatches 46; Indels 3; Gaps 2;

OY 54 DNDDETVKALASKLPSE--VFDEPDSDEIK-PQAAIIDEQAQFLTTQTAFLVAGHDE 110
 DB 68 DSKEKQYKSSQVAAFRNITPATDSYTIKGEENALILNVLWKKNNKATLYIEGHDE 127
 OY 111 RGSREYNSIGERRAVAVRNYLLGKGINQASVEIISPEERPIAFGTNEBAMSQNRRAE 169
 DB 128 RGAASYNIALGARRAAIKEHLRKQGISADRLSTISYKGEHPLNSGNHLMWQONRRTE 186

RESULT 27

092ST9

ID 092ST9 PRELIMINARY; PRT; 221 AA.

AC 092ST9;

DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

GN Hypothetical transmembrane protein SMC00354.

GN R00274 OR SMC00354.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Sinorhizobium.

OX NCBI_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,

RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

RA Godtke T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,

RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,

RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallibert F.,

RT "Analysis of the chromosome sequence of the legume symbiont

RL Sinorhizobium meliloti strain 1021";

RN Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).

DR EMBL: AL591783; CAC41711.1; -

DR InterPro: IPR001145; Bac_OmpA.

DR Pfam: PF00691; OmpA; 1.

DR ProDom: PD000930; Bac_OmpA; 1.

DR PROSITE: PS01068; OMPA; UNKNOWN.1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 221 AA; 22852 MW; 3FF7CBA8F69DD36 CRC64;

Query Match 23.7%; Score 204; DB 16; Length 221;
 Best Local Similarity 37.6%; Pred. No. 4.4e-11;
 Matches 47; Conservative 17; Mismatches 57; Indels 4; Gaps 1;

OY 47 TGVAPLVNDDETVKALASKLPSEVFDEPDSDEIKPQAAIIDEQAQFLTTQTAFLVAG 106
 DB 94 TGVSVTRAGDRIT---LNMPSNTTFATDRQVAPAFYSTLDSVAIVLRFRNKTLIDVDG 149
 OY 107 HTDERGSRREYNSIGERRAVAVRNYLLGKGINQASVEIISPEERPIAFGTNEBAMSQNR 166
 DB 150 HTDSTGASVYNGSLERRAASVANYLASRGVDQRRMAVGYGLERPIASNAATEGVRQNR 209
 OY 167 RAELS 171
 DB 210 RVEIS 214

RESULT 28

098NA9

ID 098NA9 PRELIMINARY; PRT; 216 AA.

AC 098NA9;

DT 01-OCT-2001 (TREMBlrel. 18, Created)

DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

GN Hypothetical protein m10220.

GN ME00220.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Phyllobacteriaceae; Mesorhizobium.

OX NCBI_TaxID=381;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MAFF303099;

RX MEDLINE=21082930; PubMed=11214968;

RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,

RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,

RA Takeuchi C., Yamada W., Tabata S.,

RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RT Mesorhizobium loti.
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AF002994; BAB47852.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA.1.
 DR Hypothetical protein: complete
 KW SEQUENCE 216 AA; 22401 MW; 6FC9E5B8BAC404A CRC64;

Query Match 23.5%; Score 202; DB 16; Length 216;
 Best Local Similarity 36.3%; Pred. No. 6.5e-11;
 Matches 45; Conservative 20; Mismatches 55; Indels 4; Gaps 1;

OY 47 TGVAPLVNDDETVALASKLPFLVFPDSEIKPQAAAILDEQAQFLTTNQTARVYAG 106
 DB 91 TGVSVTRSGDQII-----LNMPSDITFENVDDAVKRGEPYVLSVALYKKRQTTVVF 146
 OY 107 HTDEGRSREYNYSIGERRAVVRNVLGKGINQSVETISFGERPIAFGTNEBWSQNR 166
 DB 147 HTDSTGSDQHNFDSLQRALAVANTLSQGYDQRRFAYTGKTRPIASNTAAGREONR 206
 OY 167 RAEL 170
 DB 207 RVEI 210

RESULT 29
 OYVR19 PRELIMINARY; PRT; 249 AA.

AC 08VR19;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical 26.3 kDa protein.
 OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
 OX NCBI_Taxid=34;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kirby J.R., Zusman D.R.,
 RT "A Chemotaxis System from Myxococcus xanthus Regulates Developmental
 Gene Expression Rather than Motility".
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF448145; AAL46989.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA.1.
 KW Hypothetical protein.
 SO SEQUENCE 249 AA; 26335 MW; 170AF8EFAF632855 CRC64;

Query Match 23.4%; Score 201.5; DB 2; Length 249;
 Best Local Similarity 40.2%; Pred. No. 8.8e-11;
 Matches 41; Conservative 19; Mismatches 41; Indels 1; Gaps 1;

OY 68 SLVYFPDSDDEIKPQAAAILDEQAQFLTTNQTARVYAGHTDEGRSREYNYSIGERRAVA 127
 DB 146 SPVPFGFNSGLSADNQORSLDAQCMKT-APGRVYLAHADRGEEENLDSNRRAAS 204
 OY 128 VRNVLGKGINQSVETISFGERPIAFGTNEBWSQNR 169
 DB 205 VKRYITDLGVAPASOLGTGVGERTPVNSASSEDASSENRRVE 246

RESULT 30
 ID Q46099 PRELIMINARY; PRT; 165 AA.
 AC Q46099;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

DE OMP18 protein.
 GN PAL OR C1AD.
 OS Campylobacter jejuni.
 OG Plasmid pUM97.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_Taxid=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M275.
 RX MEDLINE=96201603; PubMed=8613402;
 RA Konkel M.E., Mead D.J., Clepiak W. Jr.,
 RT "Cloning, sequencing, and expression of a gene from Campylobacter
 jejuni encoding a protein (Omp18) with similarity to peptidoglycan-
 associated lipoproteins".
 RL Infect. Immun. 64:1850-1853(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-72DZ/92;
 RA Pawelec D.P., Jaguszczyn-Krynicka K.E.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U47617; AAC35420.1;
 DR EMBL: AJ132802; CAA10786.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA.1.
 KW Plasmid: Signal.
 SO SEQUENCE 165 AA; 17890 MW; 35E67427864B9C9D CRC64;

Query Match 23.2%; Score 199.5; DB 2; Length 165;
 Best Local Similarity 33.1%; Pred. No. 7.6e-11;
 Matches 57; Conservative 26; Mismatches 70; Indels 19; Gaps 5;

OY 2 MLHQTAAALSVLTFMTGKANKSTQVMPNAPGTGYGVY--TGVAPLVNDDET 58
 DB 4 ILSSIAAFPLVYS-----GCSTKTS-----VSGTSDVSRGTGSGDGMIDSK 49
 OY 59 VKALASKLPFLVFPDSDDEIKPQAAAILDEQAQFLTTNQTARVYAGHTDEGRSREN 117
 DB 50 ISQINDTL-GKVFYDFEFKFNIRPDMQVNTNINNEVSGVITVEGCDMGIDEN 108
 OY 118 MSIGERRAVVRNVLGKGINQSVETISFGERPIAFGTNEBWSQNR 169
 DB 109 QALGLKRAKAVKEALIQVNSDRIAVKSGETINPCTERTKACDAQNRRAE 160

RESULT 31
 ID O84605 PRELIMINARY; PRT; 188 AA.
 AC O84605;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Peptidoglycan-associated lipoprotein.
 GN PAL OR C1600.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_Taxid=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/UW-3/CX;
 RX MEDLINE=99008089; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.,
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis".
 RL Science 282:754-759(1998).
 DR EMBL: AE001330; AAC68202.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR ProDom: PD000930; Bac_OmpA.1.

KM Lipoprotein; Complete proteome.
 SQ SEQUENCE 188 AA; 21532 MW; 1A999E4AC5C31253 CRC64;
 Query Match 22.8%; Score 196.5; DB 16; Length 188;
 Best Local Similarity 35.3%; Pred. No. 1.7e-10;
 Matches 49; Conservative 19; Mismatches 54; Indels 17; Gaps 2;

OY 48 GVAPLVNDETVKALSKLPSTL-----VYFDFSDSEIKPQ-AAAILDEQ 90
 Db 44 GFVFPYSDEEIOQAFVDFDSKEQLYKTSQOSTSPRNTTFATDSYIKGEDNLTILASL 103
 OY 91 AQLFTTQARVAVLAVAGTDERGSRKYMISGERRAVAVRYNLGKGINQASVELISFGE 150
 Db 104 VRHLKSPKATLYIEGHTDERGAAYVNLGARRANAVKQYLKQGIADRLFTISYKRE 163
 OY 151 RPIAFGTNEAMSONRRAE 169
 Db 164 HPVHSGHNELAMQONRRTE 182

RESULT 32
 Q9PJEO PRELIMINARY; PRT; 202 AA.
 AC Q9PJEO
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 GN Peptidoglycan associated lipoprotein, putative.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaeae; Chlamydia.
 OX NCBI_TaxID=83360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPN / NIGG;
 RX MEDLINE=20150255; Pubmed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Knout H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE002355; AAF39684.1; -;
 DR TIGR: TC0889; -;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Complete proteome.
 SQ SEQUENCE 202 AA; 23208 MW; 02F2F13F6CDF6901 CRC64;

Query Match 22.8%; Score 196.5; DB 16; Length 202;
 Best Local Similarity 34.5%; Pred. No. 1.9e-10;
 Matches 48; Conservative 19; Mismatches 55; Indels 17; Gaps 2;

OY 48 GVAPLVNDETVKALSKLPSTL-----VYFDFSDSEIKPQ-AAAILDEQ 90
 Db 58 GFVFPYSDEEIOQAFVDFDSKEQLYKTSQOSTSPRNTTFATDSYIKGEDNLTILASL 117
 OY 91 AQLFTTQARVAVLAVAGTDERGSRKYMISGERRAVAVRYNLGKGINQASVELISFGE 150
 Db 118 VRHLKSPKATLYIEGHTDERGAAYVNLGARRANAVKQYLKQGIADRLFTISYKRE 177
 OY 151 RPIAFGTNEAMSONRRAE 169
 Db 178 HPVHSGHNELAMQONRRTE 196

RESULT 33
 Q9XCZ6 PRELIMINARY; PRT; 321 AA.
 ID Q9XCZ6

AC Q9XCZ6;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Outer membrane protein Oprf (Fragment).
 GN OPRF.
 OS Pseudomonas stutzeri A15.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=91504;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LMG 10652;
 RX MEDLINE=99319345; Pubmed=10390872;
 RA Verminten H., Willems A., Schoofs G., de Mot R., Keljers V., Hai W.,
 RA Vanderleyden J.;
 RT "The rice inoculant strain Alcaligenes faecalis A15 is a nitrogen-
 RT fixing Pseudomonas stutzeri.";
 RL Syst. Appl. Microbiol. 22:215-224(1999).
 DR EMBL: AF117972; AAD39367.1; -;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; UNKNOWN_1.
 FT NON_TER 321
 SQ SEQUENCE 321 AA; 34767 MW; 51110919C3E4FDE1 CRC64;

Query Match 22.5%; Score 193.5; DB 2; Length 321;
 Best Local Similarity 35.5%; Pred. No. 7e-10;
 Matches 43; Conservative 21; Mismatches 56; Indels 1; Gaps 1;

OY 50 APLVNDETVKALSKLPSTLYFDFSDSEIKPQAAAILDEQAQLTTQARVAVAGHTD 109
 Db 200 APLVNDSEPAPEVVRVLDVDFDKSRVSEESYSDKNADPMQYPTTVEGHTD 259
 OY 110 EGRSREYMSIGERRAVAVRYNLGK-GINQASVELISFGEPRPIAFGTNEAMSONRR 168
 Db 260 SVGTQYQYKRSERRAEAVRYNLVNEYGQNRVNSVGYGSRPVADNSTEGRQINRRV 319
 OY 169 E 169
 Db 320 E 320

RESULT 34
 Q9XCZ3 PRELIMINARY; PRT; 321 AA.
 ID Q9XCZ3
 AC Q9XCZ3;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Outer membrane protein Oprf (Fragment).
 GN OPRF.
 OS Pseudomonas stutzeri A15.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=96564;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 14405;
 RX MEDLINE=99319345; Pubmed=10390872;
 RA Verminten H., Willems A., Schoofs G., de Mot R., Keljers V., Hai W.,
 RA Vanderleyden J.;
 RT "The rice inoculant strain Alcaligenes faecalis A15 is a nitrogen-
 RT fixing Pseudomonas stutzeri.";
 RL Syst. Appl. Microbiol. 22:215-224(1999).
 DR EMBL: AF117975; AAD39370.1; -;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; UNKNOWN_1.

[illegible]

Query Match	Best Local Similarity	Matches	Score 191;	DB 2;	Length 323;
33	VAPNAPAGYGVVLTGAPLVNDDETVKALAKLEPSLVYFPDSDSEIKPQAAAILIDBOAQ	47;	Conservative 21;	Mismatches 64;	Indels 6; Gaps 2
190	VAPVEPVP-----EPAPAPVIVDEPPEAPAEVYRVELDKVPFEDKSRVNEESYDIKNLAD				
93	FLTTNQTARVLVAGHTDGRSGREYKNSLIGERRAVALVRYNLTGK-GINQASVEIITSFGEER				
245	FMQGYQPTTIVVGHGHTDSVGTQDYNQRLSERRAEAVRVNLVNEGYGQGNRVNSVGYGSESR				
152	PIAGTNEEAMSONRRAE				
305	PVADNSTEGROJNRVE				

```

AC Q9H8U8.2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE Probable outer membrane protein.
GN PA3692.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_Taxid=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Ervin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.L., Gollery L., Tolentino E., Westbrock-Madman S., Yuan Y.,
RA Brody L.L., Coulter S.H., Folger K.R., Kas A., Labib K., Lim R.M.,
RA Smith K.A., Spencer D.N., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL, AE004788; AAC07080.1; -.
DR InterPro; IPR001145; Bac_OmpA.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PRO1021; OMPADOMAIN.
DR ProDom; PD000930; Bac_OmpA; 1.
DR PROSITE; PS01068; OMPA; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 28514 MW; FA9E5F1000BDC235 CRC64;

```

Query Match	Similarity	22.0%	Score 189.5;	DB 16;	Length 261;
Best Local	Similarity	31.7%	Pred. No. 1,2e-09;		
Matches	53;	Conservative	23;	Mismatches	68;
				Indels	23;
				Gaps	2
Qy	5	10IAAAAAALSVLTMTGCAKSTQVMVAPNAPGYTGVIYGYAPLVNDNEYKALAS	64		
Db	102	IQNANSAQRAQRLDARTAQDLRSLQ-----NAKQTSRGTYTF-----	141		
Qy	65	KLPISLYPDPDSDELKPKQAALLDQAOGLTTNQTARVLYVAGHTDERSREYNLSGERR	124		
Db	142	--GVLPELDKSDLKPGKAMNIQQLAEFLQONPERQVYVESTITSTGSANTYNNRLSRR	198		
Qy	125	AAVAVNYLLGKGINQASVEIISFGGERPIAFGTNEAWSQNNRAELS	171		
Db	199	ADSVAMALLSRGISPERVATRGYKGEYPAASNGTSSGRAMNRVETV	245		
RESULT 37					
Q8UBX6					
ID	Q8UBX6	PRELIMINARY;	PRT;	220	AA.
AC	Q8UBX6				
DT	01-JUN-2002 (TREMblrel. 21, Created)				
DT	01-JUN-2002 (TREMblrel. 21, Last sequence update)				
DT	01-JUN-2002 (TREMblrel. 21, Last annotation update)				
DE	Porin.				
GN	OMPA OR ATU2722 OR AGR_C_4934.				
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
OC	Rhizobiaceae; Rhizobium.				
OX	NCBI_TaxID=176299;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21608550; PubMed=11743193;				
RA	Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,				
RA	Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,				
RA	Chen Y., Paulsen I.T., Elsen J.A., Katp P.D., Boyee D. Sr.,				
RA	Chapman P., Clendinning J., Deatherage G., Gilliet W., Grant C.,				
RA	Kutyavin T., Levy R., Li M.-J., McCelllland E., Palmeri A.,				
RA	Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,				
RA	Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,				
RA	Gordon Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,				

RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Homiel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
 RA Mollan C., Allinger M., Doughly D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,
 RA Cleto C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL: AE009218; AAL43703.1;
 DR EMBL: AE008185; AAK8440.1;
 KW Complete proteome.
 SO SEQUENCE 220 AA; 22640 MW; C3CED820F0424E2 CRC64;
 Query Match 22.0%; Score 189; DB 16; Length 220;
 Best Local Similarity 33.9%; Pred. No. 1,1e-09;
 Matches 42; Conservative 22; Mismatches 56; Indels 4; Gaps 1;
 QY 47 TGVAPLVNDETVKALASKLPVLYFDPDSDEIKPQAAAILDECAQFLTTNQTARVLYAG 106
 DB 94 TGVSVTRRGD---SIVLNMPSNITFATDQDVIPFQTLDSVAIVLNKRNRLIIDD 149
 QY 107 HTDERGSEYNNMSLGRRAVAVRNLLGKIQASVEIISGEERPIAFGTNEAWSONR 166
 DB 150 HTDSTSLQHNQALSRRAASVANYLGARGVDORISTLGGPSPPIASNATSDGRONR 209
 QY 167 RAEL 170
 DB 210 RVEV 213
 RESULT 38
 ID Q8XRSO PRELIMINARY; PRT; 277 AA.
 AC Q8XRSO;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Probable lipoprotein.
 GN RSP0761 OR RS01947.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Bacteria;
 OC Ralstonia.
 NCBI_Taxid=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM1000;
 RC MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brothier P., Camus J.C., Catolico L.,
 RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigler P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Welzenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
 RL Nature 415:497-502(2002).
 DR EMBL: AL646080; CAB17912.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Plasmid; Complete proteome.
 SO SEQUENCE 277 AA; 28898 MW; D08C29F4498FF525 CRC64;

Query Match 22.0%; Score 189; DB 16; Length 277;
 Best Local Similarity 31.3%; Pred. No. 1,5e-09;
 Matches 47; Conservative 29; Mismatches 64; Indels 10; Gaps 3;
 QY 25 NKSTQVMAPNAPPTGTYITR-----GVAPLVNDETVKALASKLPVLYFDPDSDEI 79
 DB 128 NEAQRQOLASDMATT--LGASYIKNGKLVAAVSEQSLDQTLNR---ITFEFGSNTL 182
 QY 80 KPOAAALIDQAOFLTTNQTARVLYVGHDERGSEYNNMSLGRRAVAVRNLLGKIQ 139
 DB 183 TPGRGTLDDMAVLRNLGRKLEIYGHDSGSRAMNLSQAAETVKYLLSKGAEP 242
 QY 140 ASVEIISGEERPIAFGTNEAWSONRAE 169
 DB 243 GTLTAIGVDPQVPVAPNATDEGRSRRRIE 272
 RESULT 39
 ID Q8YHLL PRELIMINARY; PRT; 232 AA.
 AC Q8YHLL;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Outer membrane protein.
 GN BMEI0786.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 NCBI_Taxid=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RC MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujar C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyprides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL: AE009519; AAL51967.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Complete proteome.
 SO SEQUENCE 232 AA; 24422 MW; 528DE7B4367E73FF CRC64;
 Query Match 21.8%; Score 187.5; DB 16; Length 232;
 Best Local Similarity 27.1%; Pred. No. 1,6e-09;
 Matches 58; Conservative 25; Mismatches 84; Indels 47; Gaps 5;
 QY 4 HIOIAAAAALSVLFMTGC-ANKSTQVAV-----APNAPTGYTGTYIG----- 48
 DB 11 HFMUKTGIALICATFLAGCTTDPYTGQKVSMTGCAIGAVALGLMGSSRAOR 70
 QY 49 -----VAPLVNDET-VKA-----DASKLPVLYFDPDS 76
 DB 71 NAVLIGAGIGALGGGLIGNVMDQEAELRAQLOGTGVSVTRNDRITLMMPSNITFDQD 130
 QY 77 DEIKPQAAALIDQAOFLTTNQTARVLYVGHDERGSEYNNMSLGRRAVAVRNLLGKIQ 136
 DB 131 DVQVSKQFYPLTNSVAIVLRKFDKTLVDYIGFTDSTGSASTNOLSRRAASVASYLDSQG 190
 QY 137 INQASVEIISGEERPIAFGTNEAWSONRAEL 170
 DB 191 IDPRRFVAVIGYGASQPIASNAETPEGRAONRIVEI 224
 RESULT 40
 Q9XCZ7

AC	09XCZ7;	PRELIMINARY;	PRT;	328 AA.
DT	01-NOV-1999 (TrEMBLrel. 12, Created)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	Outer membrane protein OprF (fragment).			
CN	OPRF.			
OS	<i>Pseudomonas stutzeri</i> (<i>Pseudomonas perfectomarina</i>).			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	<i>Pseudomonas</i> .			
OX	NCBI_TaxID=316;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=LMG 11199T.			
RX	MEDLINE=99319345; PubMed=10390872;			
RA	Vermelren H., Willems A., Schoofs G., de Mot R., Keijfers V., Hai W.,			
RA	Vanderleyden J.;			
RT	"The Rice inoculant strain Alcalligenes faecalis A15 is a nitrogen-			
RL	fixing <i>Pseudomonas stutzeri</i> .";			
RL	Syst. Appl. Microbiol. 22:215-224(1999).			
DR	EMBL; AF117971; AAD39366.1; "			
DR	InterPro: IPR001145; Bac.OmpA.			
DR	InterPro: IPR001035; MotC.			
DR	Pfam: PF00691; OmpA; 1.			
DR	PRINTS; PRO1023; NAFUGMOTY.			
DR	PRINTS; PRO1021; OMPADOMAIN.			
DR	ProDom: PD000930; Bac.OmpA; 1.			
DR	PROSITE; PS01068; OMPA; UNKNOWN_1.			
FT	NON_TER	328		
FT	SEQUENCE	328 AA; 35302 MW; 5FCDDDF79C6955090 CMC64;		

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:06:35 ; Search time 35 Seconds
(without alignments)
144.593 Million cell updates/sec

Title: US-09-674-779b-2
Perfect score: 861
Sequence: 1 MMLHIQIAAAALSVLFEM.....IAFTNEEAMQNRRAELSY 172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/1/laa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/laa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/laa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/laa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/laa/PCRTUS.COMB.pep:*
6: /cgn2_6/prodata/1/laa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	30.7	153	6	5173294-2
2	175.5	20.4	214	2	US-08-572-447C-13
3	175.5	20.4	214	4	US-09-267-747-13
4	174.5	20.3	338	1	US-08-210-394-1
5	171.5	19.9	161	2	US-08-572-447C-11
6	171.5	19.9	161	4	US-09-267-747-11
7	171.5	19.9	226	4	US-08-572-447C-15
8	171.5	19.9	226	4	US-09-267-747-15
9	165.5	19.2	359	1	US-08-457-997B-2
10	165.5	19.2	359	3	US-08-467-722A-2
11	136.5	15.9	335	4	US-08-836-500A-2
12	87.5	10.2	507	4	US-08-469-260A-605
13	87.5	10.2	1422	4	US-08-469-260A-83
14	87.5	10.2	2864	4	US-08-469-260A-394
15	84	9.8	30	1	US-08-475-989-41
16	84	9.8	30	3	US-08-475-989-41
17	84	9.8	30	3	US-08-475-989-41
18	81	9.4	964	4	US-09-556-877-177
19	81	9.4	964	4	US-09-620-412C-191
20	81	9.4	977	4	US-09-556-877-191
21	81	9.4	977	4	US-09-620-412C-191
22	79	9.2	24	1	US-08-475-985-39
23	79	9.2	24	2	US-08-475-985-39
24	79	9.2	24	2	US-08-256-839-39
25	79	9.2	2616	6	5206163-3
26	76	8.8	885	3	US-09-074-579-5
27	76	8.8	885	4	US-09-388-774-5

TITLE OF INVENTION: Immunogenic Hybrid Protein opf-opr1
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-267-747-11

Query Match 19.9%; Score 171.5; DB 4; Length 161;
Best Local Similarity 37.6%; Pred. No. 7.3e-13;
Matches 38; Conservative 15; Mismatches 47; Indels 1; Gaps 1;

QY 70 VYFDPDSDEIKPQAALILDEQAQFLTNTQARVLVAGHTDEGSRRYNMSLGERRAVAVR 129
| | | | : : : : : : : : | | | | : : | | | | | |
DB 53 VKFDFSKVKENSVDIKNLADPMKQYPTSTVEGHTDSVGDYVNOKLSERRANAVR 112
:
QY 130 NYLLGK-GINQASVEIISFGEERPIAFGTNEEAMSONRRAE 169
:
DB 113 DVLVNEGVGEGRVNAVGYGESRPVADNATAEGRALNRVE 153

RESULT 7
US-08-572-447C-15
Sequence 15, Application US/08572447C
Patent No. 5955090
GENERAL INFORMATION:
APPLICANT: Knapp, Bernhard
APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Broker, Michael
APPLICANT: Von Specht, Bernd-Ulrich
APPLICANT: Domdey, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Opt-OptI
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/572.447C
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-572-447C-15

Query Match 19.9%; Score 171.5; DB 2; Length 226;
Best Local Similarity 37.6%; Pred. No. 1.2e-12;
Matches 38; Conservative 15; Mismatches 47; Indels 1; Gaps 1;

QY 70 VYFDPDSDEIKPQAALILDEQAQFLTNTQARVLVAGHTDEGSRRYNMSLGERRAVAVR 129
| | | | : : : : : : : : | | | | : : | | | | | |
DB 118 VKFDFSKVKENSVDIKNLADPMKQYPTSTVEGHTDSVGDYVNOKLSERRANAVR 177
:
QY 130 NYLLGK-GINQASVEIISFGEERPIAFGTNEEAMSONRRAE 169
:
DB 178 DVLVNEGVGEGRVNAVGYGESRPVADNATAEGRALNRVE 218

RESULT 8
US-09-267-747-15
Sequence 15, Application US/09267747
Patent No. 6300102
GENERAL INFORMATION:
APPLICANT: Knapp, Bernhard
APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Broker, Michael
APPLICANT: Von Specht, Bernd-Ulrich
APPLICANT: Domdey, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Opt-OptI
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/267.747
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/572.447
FILING DATE: 14-DEC-1995
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:

Query Match	10.2%	Score 87.5;	DB 4;	Length 1422;
Best Local Similarity	28.4%	Pred. NO. 0.31;		
Matches	38;	Conservative 17;	Mismatches 48;	Indels 31;
			Gaps	6;
QY	3	LHIQIAAAALSVLEFMTGCAKSTSOVMAPAPAPGTYGTYTGAAPLVNDKVKAL	62	
		: : : : : : : : : : : : : :		
Db	562	LAVGVAMAAVLAIDTFGATCYRCKWSTSY----PGAL-----VAAPVDEELVEEC	611	
QY	63	ASKSLPVLVEFDESDSEIKPQAAAILDEQAOELTTNOTARLVAGHTDERSGREYNSLGE	122	
		: : : : : : : : : : : : :		
Db	612	ASRIP-----LEAMVAAL-DKLRKSTTT-----TSPFLLEALAKLNFLGSP	652	
QY	123	RRR--VAARNYLLG	134	
		: : :		
Db	653	HAATIAAILEYCCG	666	

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

QY 3 LHIQIAAAALSVLTFTMGANKSTSCVMAWAPNAPFGYITGVADLVNDETVKAL 62
I : - : I : - : I : - : I : - : I : - : I : - : I : - : I : - :
Db 1566 LANGVAVNAAYLAIDTFCATCVRKWSITSV----PLGAI-----VAIVDEBEELVEBC 1615

QY 63 ASKLBPVYFEDSDSEIRPOAAAILDEOAEFLTINOTARVLVAGHTEDEGRSREYNKSJGE 122
I : - : I : - : I : - : I : - : I : - : I : - : I : - : I : - :
Db 1616 AS-IP-----LEAMWAAL-DKLKSYTF-----TSPFLTELALKLNFPLGP 1656

QY 123 RRA--VAVERNLLG 134
I : - : I : - : I : - : I : - : I : - : I : - : I : - : I : - :
Db 1657 HATIALILEYC CG 1670

RESULT 15
US-08-475-989-41
Sequence 41, Application US/08475989
Patent No. 5679352
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: KANDILU, Ali
APPLICANT: STA, Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: Synthetic Haemophilus Influenzae
TITLE OF INVENTION: Conjugate Vaccine
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
City: Toronto
STATE: Ontario
COUNTRY: Canada
Zip: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,989
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,839
FILING DATE: 03-FEB-1993

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/CA93/00041
FILING DATE: 03-FEB-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9202219.3
FILING DATE: 03-FEB-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-506 MIS:Vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-475-989-41

Query Match 9.8%; Score 84; DB 1; Length 30;
Best Local Similarity 53.6%; Pred. No. 0.0021;
Matches 15; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 145 ISFGERRPIAGTNEAWSQNRRAELSY 172
:||||| | :|:|:|:|:|:|:|
DB 3 VSYGEEKPAVLGHDEAAYSKNRRAVLAY 30

RESULT 16
US-08-475-985-41
Sequence 41, Application US/08475985
Patent No. 5972349
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: KANDIL, Ali
APPLICANT: SIA, Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: Synthetic Haemophilus Influenzae
TITLE OF INVENTION: Conjugate Vaccine
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,985
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,839
FILING DATE: 03-FEB-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/CA93/00041
FILING DATE: 03-FEB-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9202219.3
FILING DATE: 03-FEB-1992
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-506 MIS:Vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-475-985-41

Query Match 9.8%; Score 84; DB 2; Length 30;
Best Local Similarity 53.6%; Pred. No. 0.0021;
Matches 15; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 145 ISFGERRPIAGTNEAWSQNRRAELSY 172
:||||| | :|:|:|:|:|:|:|
DB 3 VSYGEEKPAVLGHDEAAYSKNRRAVLAY 30

RESULT 17
US-08-256-839-41
Sequence 41, Application US/08256839
Patent No. 6018019
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: KANDIL, Ali
APPLICANT: SIA, Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: Synthetic Haemophilus Influenzae
TITLE OF INVENTION: Conjugate Vaccine
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,839
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-256-839-41

Query Match 9.8%; Score 84; DB 3; Length 30;
Best Local Similarity 53.6%; Pred. No. 0.0021;
Matches 15; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 145 ISFGERRPIAGTNEAWSQNRRAELSY 172
:||||| | :|:|:|:|:|:|:|

Db 3 VSYGEEKPAVLGHDEAAYSKNRRAYLAY 30

RESULT 18

US-09-556-877-177

; Sequence 177, Application US/09556877
; Patent No. 6432916
; GENERAL INFORMATION:
; APPLICANT: Probst, Peter
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skelky, Yasir
; APPLICANT: Fling, Steve
; APPLICANT: Maisonneuve, Jeff
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C5
; CURRENT APPLICATION NUMBER: US/09/556,877
; CURRENT FILING DATE: 2000-04-19
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; NUMBER OF SEQ ID NOS: 305
; SEQ ID NO 177
; LENGTH: 964
; TYPE: PRF
; ORGANISM: Chlamydia
US-09-556-877-177

Query Match 9.4%; Score 81; DB 4; Length 964;

Best Local Similarity 23.6%; Pred. No. 1;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLTFMTGCAKSTSOVWVAPNAPGTGYTGVYGVAPLVNDDET--VKALASK 65

Db 277 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNYGAIYAPVYTLVDNGPTFFINNIA 334

QY 66 LPSLYVDFDSDDEIKPQAAAILDEQAQFLTTNQTARVLVAGHTDERSREYNMGLGERRA 125

Db 335 KGAAYIDGTSN-----SKISADRAHAIIFNENIYTVTNANGTSTSNAP-----RNA 383

QY 126 VAVRN-----YLLGKGINO-----ASVEITSGEERPIAFGTNEEA 161

Db 384 ITVASSSGEILLGAGSSONLIFYDPIEVSNAG-----VVSFNKRA 424

RESULT 19

US-09-620-412C-177

; Sequence 177, Application US/09620412C

; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 177

; LENGTH: 964

; TYPE: PRF

; ORGANISM: Chlamydia

US-09-620-412C-177

Query Match 9.4%; Score 81; DB 4; Length 964;

Best Local Similarity 23.6%; Pred. No. 1;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLTFMTGCAKSTSOVWVAPNAPGTGYTGVYGVAPLVNDDET--VKALASK 65

Db 277 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNYGAIYAPVYTLVDNGPTFFINNIA 334

QY 66 LPSLYVDFDSDDEIKPQAAAILDEQAQFLTTNQTARVLVAGHTDERSREYNMGLGERRA 125

Db 335 KGAAYIDGTSN-----SKISADRAHAIIFNENIYTVTNANGTSTSNAP-----RNA 383

QY 126 VAVRN-----YLLGKGINO-----ASVEITSGEERPIAFGTNEEA 161

Db 384 ITVASSSGEILLGAGSSONLIFYDPIEVSNAG-----VVSFNKRA 424

RESULT 20

US-09-556-877-191

; Sequence 191, Application US/09556877

; Patent No. 6432916

; GENERAL INFORMATION:

; APPLICANT: Probst, Peter

; APPLICANT: Bhatia, Ajay

; APPLICANT: Skelky, Yasir

; APPLICANT: Fling, Steve

; APPLICANT: Maisonneuve, Jeff

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C5

; CURRENT APPLICATION NUMBER: US/09/556,877

; CURRENT FILING DATE: 2000-04-19

; NUMBER OF SEQ ID NOS: 305

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 191

; LENGTH: 977

; TYPE: PRF

; ORGANISM: Chlamydia

US-09-556-877-191

Query Match 9.4%; Score 81; DB 4; Length 977;

Best Local Similarity 23.6%; Pred. No. 1.1;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLTFMTGCAKSTSOVWVAPNAPGTGYTGVYGVAPLVNDDET--VKALASK 65

Db 290 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNYGAIYAPVYTLVDNGPTFFINNIA 347

QY 66 LPSLYVDFDSDDEIKPQAAAILDEQAQFLTTNQTARVLVAGHTDERSREYNMGLGERRA 125

Db 348 KGAAYIDGTSN-----SKISADRAHAIIFNENIYTVTNANGTSTSNAP-----RNA 396

QY 126 VAVRN-----YLLGKGINO-----ASVEITSGEERPIAFGTNEEA 161

Db 397 ITVASSSGEILLGAGSSONLIFYDPIEVSNAG-----VVSFNKRA 437

RESULT 21

US-09-620-412C-191

; Sequence 191, Application US/09620412C

; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 191

; LENGTH: 977

; TYPE: PRF

; ORGANISM: Chlamydia

US-09-620-412C-191

Query Match 9.4%; Score 81; DB 4; Length 977;

Best Local Similarity 23.6%; Pred. No. 1.1;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLTFMTGCAKSTSOVWVAPNAPGTGYTGVYGVAPLVNDDET--VKALASK 65

Db 290 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNYGAIYAPVYTLVDNGPTFFINNIA 347

QY 66 LPSLVYEDFSDDEIKPQAAAILDQAOFLTNTQARVLVAGHTDGRSREYNMSLGERRA 125
 DB 348 KGAAYIDGTSN-----SKISADRRALIFENENITVNTNAGSTSNMP-----RRNA 396
 QY 126 VAVR-----YLLGKING-----ASVELISGGEERPIAFGNEEA 161
 DB 397 ITVASSGEIILGAGSSONLIFYDPIEVSNA-----VSVSNKEA 437

RESULT 22
 US-08-475-989-39

; Sequence 39, Application US/08475989
 ; Patent No. 5679352
 ; GENERAL INFORMATION:
 ; APPLICANT: CHONG, Pele
 ; APPLICANT: KANDIL, Ali
 ; APPLICANT: SIA, Charles
 ; APPLICANT: KLEIN, Michel
 ; TITLE OF INVENTION: Synthetic Haemophilus Influenzae
 ; NUMBER OF SEQUENCES: 56
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/475,989
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/256,839
 ; FILING DATE: 03-FEB-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/CA93/00041
 ; FILING DATE: 03-FEB-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9202219.3
 ; FILING DATE: 03-FEB-1992
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, MICHAEL I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-505 MIS:Vg
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 39:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 24 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-475-989-39

Query Match 9.2%; Score 79; DB 1; Length 24;
 Best Local Similarity 62.5%; Pred. No. 0.0059;
 Matches 15; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 111 RGSREYNMSLGERRAVAVRNYLLG 134
 DB 1 RGTPEYNIALGRRADAVKGYLAG 24

RESULT 23
 US-08-475-985-39

; Sequence 39, Application US/08475985
 ; Patent No. 597349
 ; GENERAL INFORMATION:
 ; APPLICANT: CHONG, Pele
 ; APPLICANT: KANDIL, Ali
 ; APPLICANT: SIA, Charles
 ; APPLICANT: KLEIN, Michel
 ; TITLE OF INVENTION: Synthetic Haemophilus Influenzae
 ; NUMBER OF SEQUENCES: 56
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/475,985
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/256,839
 ; FILING DATE: 03-FEB-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/CA93/00041
 ; FILING DATE: 03-FEB-1993
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9202219.3
 ; FILING DATE: 03-FEB-1992
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, MICHAEL I.
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-506 MIS:Vg
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 39:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 24 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-475-985-39

Query Match 9.2%; Score 79; DB 2; Length 24;
 Best Local Similarity 62.5%; Pred. No. 0.0059;
 Matches 15; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 111 RGSREYNMSLGERRAVAVRNYLLG 134
 DB 1 RGTPEYNIALGRRADAVKGYLAG 24

RESULT 24
 US-08-236-839-39

; Sequence 39, Application US/08256839
 ; Patent No. 6018019
 ; GENERAL INFORMATION:
 ; APPLICANT: CHONG, Pele
 ; APPLICANT: KANDIL, Ali
 ; APPLICANT: SIA, Charles
 ; APPLICANT: KLEIN, Michel

TITLE OF INVENTION: Synthetic Haemophilus Influenzae
TITLE OF INVENTION: Conjugate Vaccine
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,839
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-256-839-39

Query Match 9.2%; Score 79; DB 3; Length 24;
Best Local Similarity 62.5%; Pred. No. 0.0059;
Matches 15; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 111 RGSREYMSIGERRAVALRYNYLG 134
DB 1 RGPPEYNALGRRADAVKGYLAG 24

RESULT 25
5206163-3
Patent No. 5206163
APPLICANT: RENARD, ANDRE; DINA, MARTIAL, JOSEPH
TITLE OF INVENTION: DNA ENCODING BOVINE DIARRHEA
VIRUS PROTEIN
NUMBER OF SEQUENCES: 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/550,816
FILING DATE: 06-JUL-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 331,037
FILING DATE: 29-MAR-1989
APPLICATION NUMBER: 752,981
FILING DATE: 08-JUL-1985
SEQ ID NO: 3;
LENGTH: 2616
5206163-3

Query Match 9.2%; Score 79; DB 6; Length 2616;
Best Local Similarity 21.8%; Pred. No. 8.5;
Matches 38; Conservative 28; Mismatches 68; Indels 40; Gaps 7;

QY 28 TSQVAVAPNAPGTGTVYGVADLVNDVTKALASKLSLVYFEDS--DEIKPQAA 84
DB 1675 TSEVGI---TLVGRALMTGITITVVEKTE---FNADGSPESIKIGLDEGCYDGPPODH 1728
QY 85 AILDEQAQFLTTNQTARLVAGHDDERSREYN-----MSLGERRAYA 127
DB 1729 TLADE---IHSRDERPFVLVIGSRSSMSNRAKTARNTINCITQKRPQEI RDLMAQGRMLVVA 1785

QY 128 VRNY-----LLGKGINASVEIISFGEERP--IAGTNEANSQNRRAEL 170
DB 1786 LRSFNPSELVDKGFLLDRVALRLSLSPGRKQVYATYAKVELLEQEQVFI 1839

RESULT 26
US-09-074-579-5
Sequence 5, Application US/09074579
Patent No. 6001596
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Guegler, Karl J.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,579
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerione, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0505 US
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 885 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENE BANK
CLONE: g1288563
US-09-074-579-5

Query Match 8.8%; Score 76; DB 3; Length 885;
Best Local Similarity 25.5%; Pred. No. 3.7;
Matches 14; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 71 YFDESDETKPOAAAILDEQAQFLTTNQTARLVAGHDDERSREYNMSIGERRA 125
DB 177 HFEIEVDIFEPQGISMLDAEASFINDLIGSALTRSFSGKGVSPKSLDOORS 231

RESULT 27
US-09-388-774-5
Sequence 5, Application US/09388774
Patent No. 6228991
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Guegler, Karl J.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE

RESULT 28
 US-08-762-106-9
 ; Sequence 9, Application US/08762106
 ; Patent No. 5948677
 ;
 ; GENERAL INFORMATION:
 ;
 APPLICANT: Jarvik, Jonathan W.
 TITLE OF INVENTION: READING FRAME INDEPENDENT EPIOTOPE
 TITLE OF INVENTION: TAGGING
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Harris Brotman
 STREET: 202 Coast Blvd., Suite 111
 CITY: La Jolla
 STATE: California
 COUNTRY: US
 ZIP: 92037
 ;
 ; COMPUTER READABLE FORM:
 ;
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/762,106
 FILING DATE: 09-DEC-1996

```

Query Match          8.5%; Score 73.5; DB 2; Length 527;
Best Local Similarity 21.2%; Pred. No. 3.3;
Matches 33; Conservative 21; Mismatches 77; Indels 25; Gaps 5

QY      38  PIGYTVIITGAPLVNDDEIVKALSKLSLV---YDPDSD-----EIK 80
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      59  PADATQTQTKGSPSPASTKRRTRTLPPRRPEAVDGRHIDQTVYLELTDTVPEADTS 118
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      81  PQAAILDE-QAQFLTTNOTARVAVAGHTDEGRSREYNSIGER-----RAAVANYLLG 134
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      119  TQTDALFLDRPPPLPVPQKTGTDAITQITDRQTDKGTIGENDLFDPEFEVPIILEVLVG 178
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      135  KGINASVEIISFGEERPIAFGTNEAWSONRAEL 170
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      179  KYLEQGLMEVLE--EEELAMRAHQEHFQIRNAEL 212
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 29
US-09-320-774-9
: Sequence 9, Application US/09320774
: Patent No. 6265345
: GENERAL INFORMATION:
: APPLICANT: Jarvik, Jonathan W.
: TITLE OF INVENTION: READING FRAME INDEPENDENT EPTIOTPE
: TITLE OF INVENTION: TAGGING
: NUMBER OF SEQUENCES: 47
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Harris Brotman
: STREET: 202 Coast Blvd., Suite 111
: CITY: La Jolla
: STATE: California
: COUNTRY: US
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/320,774
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/762,106
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brotman, Harris F.
: REGISTRATION NUMBER: 35,461
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 654-2428
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 527 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: US-09-320-774-9

```


Query Match 8.5%; Score 73.5; DB 4; Length 527;
Best Local Similarity 21.2%; Pred. No. 3.3;
Matches 33; Conservative 21; Mismatches 77; Indels 25; Gaps 5;

QY 38 PTGTGVTYTVAVPLVNDENVKALSKPLSLV-----YFDPDSD-----EIK 80
DB 59 PADATQTQTKGSPASTKRTTRLPPTPEAVDGRHIDQTVYLELTIDVPEADTS 118
QY 81 PQAALIDE-QAQLFTTQTAFLVAGHTDERGSREYNMSLGER-----RAVAARNYLLG 134
DB 119 TQTDALFDRPPPLPPLVPOKCTDALTQTTDRQTDROQGIENGDLDFDFEVEPILEVAVG 178
QY 135 KGINGASVEIISFGSEPIAPGTNEAMSQNRRAEL 170
DB 179 KVLQGLMEVLE--EELIAMAMRAHQEHFEQIRNAEL 212

RESULT 30
US-09-345-882-29
; Sequence 29, Application US/09345882
; Patent No. 639373
; GENERAL INFORMATION:
; APPLICANT: Bouquelere, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; TITLE OF INVENTION: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 29
; LENGTH: 1312
; TYPE: PR1
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 294..296
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 432..434
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 755..757
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 856..858
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 859..861
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 910..912
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 1151..1153
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 1226..1228
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION

LOCATION: 102..105
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 663..666
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 808..811
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 885..888
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 17..19
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 31..33
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 41..43
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 100..102
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 140..142
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 216..218
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 471..473
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 507..509
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 531..533
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 591..593
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 656..658
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 801..803
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 812..814
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 815..817
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 876..878


```

Sequence 2, Application US/08960756
Patent No. 5866422
GENERAL INFORMATION:
APPLICANT: WAYNE, JAY
APPLICANT: XU, SHUANG-YONG
TITLE OF INVENTION: METHOD FOR CLONING AND
TITLE OF INVENTION: PRODUCING THE Tsp45I RESTRICTION ENDONUCLEASE IN E. COLI
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: New England Biolabs, Inc.
STREET: 32 Tozer Road
CITY: Beverly
STATE: MA
COUNTRY: US
ZIP: 01915
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,756
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Gregory D
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-128
TELECOMMUNICATION INFORMATION:
TELEPHONE: 978-927-5054
TELEFAX: 978-927-1705
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 413 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-960-756-2

Query Match      8 4%; Score 72.5; DB 2; Length 413;
Best Local Similarity 23.6%, Pred No. 3;
Matches 34; Conservative 19; Mismatches 72; Indels 19; Gaps 5;

Oy    38 PGTGTTG---VIT-----TGAAPLVNDDETAKALSKLPISLYTFDFSDSETRKPPAAAILD 88
       ||| | : : : || | | | | : : | | | |
Db    87 PGCAAGVGLVYIDIPYTGQGFIVGGDETDHATVSQPKGOLGYDTLLDGPGVFELR 146
Oy    89 EQAOFLLTNQTARVLVAAGHTDDSGSREYMSL---GERRAVANYNYLLGKIGIQASVEI 144
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    147 ERLLLEIMADSLFVHIDEYGEYVALIIDVEFGRRNV---NHIAIASPNKNMR 203
Oy    145 ISFEGERPIAF--GTNEEAMSON 165
       :|| :: : : | : | : :
Db    204 KAFSGOKDMILVYSKTRDYWMNS 227

RESULT 34
US-07-920-281C-2
; Sequence 2, Application US/07920281C
; Patent No. 5739026
GENERAL INFORMATION:
APPLICANT: Garoff, Henrik
APPLICANT: Liljestrom, Peter
TITLE OF INVENTION: DNA Expression Systems Based on
TITLE OF INVENTION: Alphaviruses
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/920,281C
FILING DATE: 13-AUG-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-920-281C-2

Query Match 8.4%; Score 72.5; DB 1; Length 2431;
Best Local Similarity 23.0%; Pred. No. 46;
Matches 26; Conservative 20; Mismatches 36; Indels 31; Gaps 5;

QY 24 ANKSTQVAVAPNAPGTGYTGVAPLVNDENVKALASKLPSTLYFPDSDSEIKPQA 83
DB 1756 ASRAERPVAPRKPFT-----PAPRT-----AFRKLKP-LTFGDFDEHVDALA 1798
QY 84 AAI-----LDEQAQFLTNOTARVAVAGHTDGRSGREYMSIGERRAY 126
DB 1799 SGITFGDFDVLRLGRAGATYFSSDTG----SGHLQOKSVQHNLQCAQLDAV 1847

RESULT 35
US-08-466-277-2
Sequence 2, Application US/08466277
Patent No. 6190666
GENERAL INFORMATION:
APPLICANT: Garoff, Henrik
TITLE OF INVENTION: DNA Expression Systems Based on
Alphaviruses
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,277
FILING DATE: 06-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/920,281
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-466-277-2

Query Match 8.4%; Score 72.5; DB 4; Length 2431;
Best Local Similarity 23.0%; Pred. No. 46;
Matches 26; Conservative 20; Mismatches 36; Indels 31; Gaps 5;

QY 24 ANKSTQVAVAPNAPGTGYTGVAPLVNDENVKALASKLPSTLYFPDSDSEIKPQA 83
DB 1756 ASRAERPVAPRKPFT-----PAPRT-----AFRKLKP-LTFGDFDEHVDALA 1798
QY 84 AAI-----LDEQAQFLTNOTARVAVAGHTDGRSGREYMSIGERRAY 126
DB 1799 SGITFGDFDVLRLGRAGATYFSSDTG----SGHLQOKSVQHNLQCAQLDAV 1847

RESULT 36
US-08-110-786A-8
Sequence 8, Application US/08110786A
Patent No. 5443966
GENERAL INFORMATION:
APPLICANT: FAIRWEATHER, Neil Fraser
TITLE OF INVENTION: Expression of tetanus toxin fragment C
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye P.C.
STREET: 1100 No. 5443966th Gleebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/110,786A
FILING DATE: 23-AUG-1993 1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/777,337
FILING DATE: 29-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB90/00943
FILING DATE: 20-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8914122.0
FILING DATE: 20 June 1989
ATTORNEY/AGENT INFORMATION:
NAME: Maty J. Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 117-134
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 8:

COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,585
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/237,940
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27866/31954
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-854-585-2

Query Match
Best Local Similarity 21.8%; Pred. No. 49;
Matches 24; Conservative 26; Mismatches 46; Indels 14; Gaps 4;

QY 21 TGCANKSTSOVWVAPNAPGTGYITGVAPLVNDDETVKALASKLPSLVYDFDSEIK 80
Db 321 SGOQSRDTEVLVGLPGTRYNATVYSSAANGTEGP--QAIERTNAIQVFDVAVNIS 378
QY 81 PQAAILL---DEQAQFLTTNQTARVLVAGHTDERGSRXYMSLGERRAV 126
Db 379 ATSLTLIKVSDNES---SSNVTYKIHVAGETDSS---NLNVSEPRAV 420

RESULT 40
PCT-US95-05512-2
Sequence 2, Application PC/TUS9505512
GENERAL INFORMATION:
APPLICANT: Tonks, Nicholas K. and Stman, Arne
TITLE OF INVENTION: Density Enhanced Protein Tyrosine
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 233 South Wacker Drive, Suite 6300
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05512
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25,447
REFERENCE/DOCKET NUMBER: 27866/31954
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 1337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05512-2

Query Match
Best Local Similarity 21.8%; Pred. No. 49;
Matches 24; Conservative 26; Mismatches 46; Indels 14; Gaps 4;

QY 21 TGCANKSTSOVWVAPNAPGTGYITGVAPLVNDDETVKALASKLPSLVYDFDSEIK 80
Db 321 SGOQSRDTEVLVGLPGTRYNATVYSSAANGTEGP--QAIERTNAIQVFDVAVNIS 378
QY 81 PQAAILL---DEQAQFLTTNQTARVLVAGHTDERGSRXYMSLGERRAV 126
Db 379 ATSLTLIKVSDNES---SSNVTYKIHVAGETDSS---NLNVSEPRAV 420

Search completed: July 6, 2003, 14:08:12
Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:10:35 ; Search time 307 Seconds
(without alignments)
64.426 Million cell updates/sec

Title: US-09-674-779b-2
Perfect score: 861
Sequence: 1 MMHIOIAAALSVLTFM.....IAGTNEAMQNRRAELSY 172

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries:

Database :

1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	204.5	23.8	192	10	US-09-747-348-2
2	176.5	20.5	353	9	US-10-203-942-9
3	149	17.3	226	9	US-10-156-761-10782
4	136.5	15.9	344	9	US-10-159-953-2
5	107	12.4	380	9	US-09-998-279-26
6	107	12.4	385	9	US-09-998-279-24
7	93.5	10.9	237	9	US-10-010-160-12
8	87.5	10.2	507	8	US-08-424-5508-605
9	87.5	10.2	1422	8	US-08-424-5508-83
10	87.5	10.2	2864	8	US-08-424-5508-394
11	87.5	10.2	2864	10	US-09-742-659-2
12	87.5	10.2	2865	10	US-09-742-659-6
13	82	9.5	871	10	US-09-886-468-21
14	81	9.4	964	10	US-09-841-132-177
15	81	9.4	977	10	US-09-841-132-191
16	80	9.3	361	12	US-10-007-693-74
17	80	9.3	420	9	US-10-156-761-9543
18	77.5	9.0	877	9	US-10-156-761-13758
19	76.5	8.9	438	10	US-09-815-242-14076

20	76	8.8	885	10	US-09-828-423-5	Sequence 5, App1
21	75.5	8.8	155	9	US-10-156-761-13516	Sequence 13516, A
22	75.5	8.8	794	10	US-09-815-242-5697	Sequence 5697, Ap
23	75.5	8.8	802	10	US-09-815-242-12668	Sequence 12668, A
24	74.5	8.7	1330	9	US-10-156-761-10510	Sequence 10510, A
25	74	8.6	1177	9	US-10-128-714-3493	Sequence 3493, Ap
26	74	8.6	1179	9	US-10-128-714-8493	Sequence 8493, Ap
27	73.5	8.5	569	9	US-09-738-626-6179	Sequence 6179, Ap
28	73.5	8.5	928	10	US-09-815-242-10417	Sequence 10417, A
29	73	8.5	1312	10	US-10-071-179-2041	Sequence 29, App1
30	73	8.5	1491	10	US-09-815-242-5568	Sequence 5568, Ap
31	73	8.5	1502	10	US-09-815-242-12162	Sequence 12162, A
32	73	8.5	1520	9	US-09-738-626-4444	Sequence 4444, Ap
33	72.5	8.4	851	9	US-10-156-761-99333	Sequence 9933, Ap
34	72.5	8.4	2431	10	US-09-901-106-2	Sequence 2, App1
35	71.5	8.3	441	9	US-10-156-761-13680	Sequence 13680, A
36	71.5	8.3	507	9	US-10-156-761-14476	Sequence 14476, A
37	71.5	8.3	2862	10	US-09-742-659-5	Sequence 5, App1
38	71	8.2	248	10	US-09-925-301-908	Sequence 908, App
39	71	8.2	259	9	US-09-996-634-133	Sequence 133, App
40	71	8.2	259	9	US-09-997-181-133	Sequence 133, App
41	71	8.2	259	9	US-09-997-182-133	Sequence 133, App
42	71	8.2	1352	9	US-10-156-761-9867	Sequence 9867, Ap
43	70.5	8.2	199	9	US-09-738-626-5191	Sequence 5191, Ap
44	70.5	8.2	275	9	US-10-156-761-12256	Sequence 12256, A
45	70.5	8.2	299	9	US-10-156-761-12949	Sequence 12949, A
46	70.5	8.2	550	9	US-09-906-419-41	Sequence 41, App1
47	70.5	8.2	550	9	US-10-119-136-41	Sequence 41, App1
48	70.5	8.2	568	9	US-10-136-960-8	Sequence 8, App1
49	70	8.1	336	10	US-09-815-242-11899	Sequence 11899, A
50	70	8.1	501	9	US-09-738-626-6084	Sequence 6084, Ap
51	69.5	8.1	402	9	US-09-712-363-147	Sequence 147, App
52	69.5	8.1	645	8	US-08-834-666A-18	Sequence 18, App1
53	69	8.0	350	9	US-10-169-046-24	Sequence 24, App1
54	69	8.0	1215	10	US-09-815-242-5908	Sequence 5908, Ap
55	69	8.0	1267	10	US-09-815-242-13113	Sequence 13113, A
56	69	8.0	2339	10	US-09-815-242-5834	Sequence 5834, Ap
57	69	8.0	6281	10	US-09-815-242-12996	Sequence 12996, A
58	68.5	8.0	463	9	US-09-816-467-2	Sequence 2, App1
59	68.5	8.0	867	9	US-10-177-744A-11	Sequence 11, App1
60	68.5	8.0	903	9	US-09-746-783-142	Sequence 142, App
61	68	7.9	389	9	US-10-156-761-13369	Sequence 13369, A
62	68	7.9	438	10	US-10-156-761-9344	Sequence 9344, Ap
63	68	7.9	568	10	US-09-815-242-5140	Sequence 5140, Ap
64	67.5	7.8	327	9	US-10-156-761-11941	Sequence 11941, A
65	67.5	7.8	327	9	US-10-156-761-13045	Sequence 13045, A
66	67.5	7.8	343	9	US-10-156-761-13124	Sequence 13124, A
67	67.5	7.8	436	10	US-09-815-242-13218	Sequence 13218, A
68	67.5	7.8	438	10	US-09-741-669-452	Sequence 452, App
69	67.5	7.8	438	10	US-09-815-242-10029	Sequence 10029, A
70	67	7.8	223	9	US-09-892-398-22	Sequence 22, App1
71	67	7.8	471	9	US-10-156-761-10494	Sequence 10494, A
72	67	7.8	750	9	US-10-156-761-12865	Sequence 12865, A
73	67	7.8	760	9	US-09-892-398-29	Sequence 29, App1
74	67	7.8	761	9	US-09-892-398-1	Sequence 1, App1
75	67	7.8	1325	9	US-10-154-452-6	Sequence 6, App1
76	67	7.8	1437	9	US-10-154-452-2	Sequence 2, App1
77	67	7.8	1437	9	US-09-934-421A-6	Sequence 6, App1
78	67	7.8	1437	9	US-10-162-012-38	Sequence 38, App1
79	66.5	7.7	132	10	US-09-925-297-779	Sequence 779, App
80	66.5	7.7	193	9	US-09-738-626-3874	Sequence 3874, Ap
81	66.5	7.7	314	9	US-10-156-761-13748	Sequence 13748, A
82	66.5	7.7	797	9	US-10-156-761-10907	Sequence 10907, A
83	66.5	7.7	1187	9	US-10-223-070-23	Sequence 23, App1
84	66	7.7	421	9	US-09-738-626-6362	Sequence 6362, Ap
85	66	7.7	423	9	US-10-156-761-12135	Sequence 12135, A
86	66	7.7	510	10	US-10-124-800-26	Sequence 26, App1
87	66	7.7	565	10	US-09-765-221-218	Sequence 218, App
88	66	7.7	998	9	US-10-101-464A-331	Sequence 931, App
89	66	7.7	1622	9	US-10-331-061-72	Sequence 72, App1
90	66	7.7	2059	9	US-10-124-800-4	Sequence 4, App1
91	66	7.7	5877	9	US-10-142-515-11	Sequence 11, App1
92	66	7.7	5935	9	US-10-243-243A-8	Sequence 8, App1

ALIGNMENTS

Query Match	23.8%;	Score 204.5;	DB 10;	Length 192;
Best Local Similarity	38.7%;	Pred. No. 6.2e-14;		
Matches 46; Conservative	24;	Mismatches 46;	Indels 3;	Gaps 2

RESULT 2

```

Sequence 9, Application US/10203942
Publication No. US20030096370A1
GENERAL INFORMATION:
APPLICANT: BERTHER, FRANCOIS-XAVIER
APPLICANT: DEMOEL, PHILIPPE
APPLICANT: POOLMAN, JAN
APPLICANT: THONNARD, JOELLE
TITLE OF INVENTION: HAEMOPHILUS INFLUENZA OUTR MEMBRANE
FILE REFERENCE: 945210
CURRENT APPLICATION NUMBER: US/10/203, 942
CURRENT FILING DATE: 2002-08-15
PRIOR APPLICATION NUMBER: PCT/EP01/01556
PRIOR FILING DATE: 2001-02-13
PRIOR APPLICATION NUMBER: GB 0003502.2
PRIOR FILING DATE: 2000-02-15
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 353
TYPE: PRF
ORGANISM: Haemophilus influenzae
US-10-203-942-9

```

Query Match 20.5%; Score 176.5; DB 9; Length 353;

```

QY      144 IISFGEERPIAFGTNEE-----AWSQNRRAELS 171
      :|: | |:: :|||::
Db      312 ATGYGKANPVTGATCDQYKGRKALIACFAPDRVEIA 348

```

ORGANISM: Streptomyces avermitilis
US-10-156-761-10782

Matches 5

Db 187 SEDYPIADNTSEQGRKRRVEVTF 211

APPLICANT: François LAMNY


```

; APPLICANT: Jean-Yves BONNEFOY
; TITLE OF INVENTION: METHOD FOR PREPARING A POLYPEPTIDE SOLUBLE IN AN AQUEOUS SOLVENT
; FILE REFERENCE: D18390
; CURRENT APPLICATION NUMBER: US/10/169,953
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: FR 00 00070
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; OTHER INFORMATION: P40
US-10-169-953-2

Query Match
Best Local Similarity 15.9%; Score 136.5; DB 9; Length 344;
Matches 35; Conservative 20; Mismatches 50; Indels 11; Gaps 2;

OY 66 LPSLYFPDSDSEIKPQAAIILDEQAQFLT--NOTARVLYAGHTDEGSRKYNMGLGER 123
DB 215 LKSDVLENNKATILPEGOQALDQLYTQLSNMPPKDSGSAVYLGYTRIGRSEAYNQOLSEK 274
OY 124 RAVAVRNYLLGKGINQASVEIISFGEEPRPIAFGTNE-----EAMSONRAEL 170
DB 275 RAOSVVDYLVAKGITRAGKISARGMGESNPVTGNTCDNKARALIDCLAPRREVT 330

RESULT 5
US-09-998-279-26
; Sequence 26, Application US/09998279
; Publication No. US20030083287A1
; GENERAL INFORMATION:
; APPLICANT: BURGESS, NICOLA A.
; APPLICANT: GARCIA, MIGUEL M.
; APPLICANT: KIRKE, DAVID F.
; APPLICANT: MEYERS, NICHOLAS L.
; APPLICANT: WILLIAMS, PAUL
; TITLE OF INVENTION: gins
; FILE REFERENCE: GMS0081
; CURRENT APPLICATION NUMBER: US/09/998,279
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,288
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-998-279-26

Query Match
Best Local Similarity 12.4%; Score 107; DB 9; Length 380;
Matches 40; Conservative 16; Mismatches 49; Indels 26; Gaps 6;

OY 35 PNAPTGTYGTYTGAAPLVNDDETVKALSKLPISLVYFPDSDSEIKPQAAIILDEQAQFL 94
DB 267 PQOPT-VTRVY-----VDN-----VVYFRINSKIDRNOENINYNTAEYA 305
OY 95 TTNOTARVLYAGHTDER-GSRKYNMGLGERRAVAVRNYLLGKGINQASVEIISFGEEPRPI 153
DB 306 KTN-NAPIKVVGYADEKGTATAYNNKLSERRAKAVAKMLEKYGVSADRITTEMKSSSEOI 364
OY 154 AFGTNEAMSQ 164
DB 365 ---YEENAMNR 372

RESULT 6
```

```

US-09-998-279-24
; Sequence 24, Application US/09998279
; Publication No. US20030083287A1
; GENERAL INFORMATION:
; APPLICANT: BURGESS, NICOLA A.
; APPLICANT: GARCIA, MIGUEL M.
; APPLICANT: KIRKE, DAVID F.
; APPLICANT: MEYERS, NICHOLAS L.
; APPLICANT: WILLIAMS, PAUL
; TITLE OF INVENTION: gins
; FILE REFERENCE: GMS0081
; CURRENT APPLICATION NUMBER: US/09/998,279
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,288
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-998-279-24

Query Match
Best Local Similarity 12.4%; Score 107; DB 9; Length 385;
Matches 40; Conservative 16; Mismatches 49; Indels 26; Gaps 6;

OY 35 PNAPTGTYGTYTGAAPLVNDDETVKALSKLPISLVYFPDSDSEIKPQAAIILDEQAQFL 94
DB 272 PQOPT-VTRVY-----VDN-----VVYFRINSKIDRNOENINYNTAEYA 310
OY 95 TTNOTARVLYAGHTDER-GSRKYNMGLGERRAVAVRNYLLGKGINQASVEIISFGEEPRPI 153
DB 311 KTN-NAPIKVVGYADEKGTATAYNNKLSERRAKAVAKMLEKYGVSADRITTEMKSSSEOI 369
OY 154 AFGTNEAMSQ 164
DB 370 ---YEENAMNR 377

RESULT 7
US-10-010-160-12
; Sequence 12, Application US/10010160
; Publication No. US20030103999A1
; GENERAL INFORMATION:
; APPLICANT: Rosey, Everett L.
; APPLICANT: Strugnell, Richard A.
; APPLICANT: Good, Robert T.
; APPLICANT: King, Kendall W.
; TITLE OF INVENTION: NOVEL THERAPEUTIC COMPOSITIONS FOR
; FILE REFERENCE: DAV110.001AUS
; CURRENT APPLICATION NUMBER: US/10/010,160
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: AU PR1381
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: US 60/249,596
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-10-010-160-12

Query Match
Best Local Similarity 10.9%; Score 93.5; DB 9; Length 237;
Matches 36; Conservative 29; Mismatches 70; Indels 19; Gaps 6;

OY 24 ANKSTQVWVAPNAPPTGTYGTYTGAAPLVNDDETVKALSKLPISLVYFPDSDSEIKPQAA 83
DB 86 ANK-LKKMLMADAIPOSATGISADVDGVL---RVNSNSTFPFGTATITLPECKKVMGTIV 140
```

OY 84 AALDEQAQFLTTNOTARVLVAGHTD-----ERGSR-EYNMSLGERRAVAVNYLLGKGIN 138
DB 141 LAVLRVYLYL-----VIRGHADIGCEITKSGSPFASNNELSGARAAQAQVLYVEHGK 192
OY 139 QASVEIISFGERRIAGCTNEAMSQNRRELST 172
DB 193 ASRIRSVGYADTRPLE-PSSEGSTKNRRIEFYF 225

RESULT 8

US-08-424-550B-605
; Sequence 605, Application US/08424550B
; Patent No. US20020119447A1
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMT J. PILOT-MATTIAS
; APPLICANT: GEORGE J. DAMSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUEHROFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUTIK
; APPLICANT: ISA K. MUSHAMMAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550B
; FILING DATE:
; CLASSIFICATION: 435435
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 605:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 507 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-424-550B-605

Query Match 10.2%; Score 87.5; DB 8; Length 507;
Best Local Similarity 28.4%; Pred. No. 0.75;
Matches 38; Conservative 17; Mismatches 48; Indels 31; Gaps 6;

OY 3 LHIOIAAAAALSVITFTMGCAKSTSQVAVNAPRTGYTYTGAVAPLYDNDEYKAL 62
DB 360 LAVGVAVMAVLAIDTFGATCVRRCWSTTSV-----PTGAT-----VAPVDEEIVEEC 409
OY 63 ASKLPSLVYFDPDSDEIRPOAAAILDEQAQFLTTNOTARVLVAGHTDEGSRREYNMSLGE 122
DB 410 ASFLP-----LEAVVAAL-DKLSKSTITT-----TSPFTLETALEKLNFTLP 450
OY 123 RRA--VAVRNYLLG 134

DB 451 HAATLTAITEYCCG 464

RESULT 9

US-08-424-550B-83
; Sequence 83, Application US/08424550B
; Patent No. US20020119447A1
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMT J. PILOT-MATTIAS
; APPLICANT: GEORGE J. DAMSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUEHROFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUTIK
; APPLICANT: ISA K. MUSHAMMAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550B
; FILING DATE:
; CLASSIFICATION: 435435
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-424-550B-83

Query Match 10.2%; Score 87.5; DB 8; Length 1422;
Best Local Similarity 28.4%; Pred. No. 3.2;
Matches 38; Conservative 17; Mismatches 48; Indels 31; Gaps 6;

OY 3 LHIOIAAAAALSVITFTMGCAKSTSQVAVNAPRTGYTYTGAVAPLYDNDEYKAL 62
DB 562 LAVGVAVMAVLAIDTFGATCVRRCWSTTSV-----PTGAT-----VAPVDEEIVEEC 611
OY 63 ASKLPSLVYFDPDSDEIRPOAAAILDEQAQFLTTNOTARVLVAGHTDEGSRREYNMSLGE 122
DB 612 ASFLP-----LEAVVAAL-DKLSKSTITT-----TSPFTLETALEKLNFTLP 652
OY 123 RRA--VAVRNYLLG 134
DB 653 HAATLTAITEYCCG 666

RESULT 10
US-08-424-550B-394

```
; Sequence 394, Application US/08424550B
; Patent No. US20020119447A1
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMU J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAMSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUEHROFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUTIK
; APPLICANT: ISA K. MUSHAWAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550B
; FILING DATE:
; CLASSIFICATION: 435435
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 394:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2864 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-424-550B-394

Query Match 10.2%; Score 87.5; DB 8; Length 2864;
Best Local Similarity 28.4%; Pred. No. 8.8;
Matches 38; Conservative 17; Mismatches 48; Indels 31; Gaps 6;

QY 3 LHIOIAAAAALSVLPMTGCAKSTQVWVAPNAPTGYTGVTGVAPLVNDDETVKAL 62
DB 1566 LAVGVGVAAMAYLAIDFGATCVRKWSITSV---PTGAT-----VAPVDEEIVEEC 1615
QY 63 ASKLPSLVYDFDSDDEIKPQAAAILDEQAQFLTTNOTARVLVAGHTDGRSGREYNMSLGE 122
DB 1616 ASFLP-----LEAMVAAL-DKLKSTITT-----TSPFLLEALEKINTFLGP 1656
QY 123 RRA--VAVRNYLLG 134
DB 1657 HAATLTAITEYCCG 1670

RESULT 11
US-09-742-659-2
; Sequence 2, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
```

```
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2864
; TYPE: PRT
; ORGANISM: GB virus-B
; US-09-742-659-2

Query Match 10.2%; Score 87.5; DB 10; Length 2864;
Best Local Similarity 28.4%; Pred. No. 8.8;
Matches 38; Conservative 17; Mismatches 48; Indels 31; Gaps 6;

QY 3 LHIOIAAAAALSVLPMTGCAKSTQVWVAPNAPTGYTGVTGVAPLVNDDETVKAL 62
DB 1566 LAVGVGVAAMAYLAIDFGATCVRKWSITSV---PTGAT-----VAPVDEEIVEEC 1615
QY 63 ASKLPSLVYDFDSDDEIKPQAAAILDEQAQFLTTNOTARVLVAGHTDGRSGREYNMSLGE 122
DB 1616 ASFLP-----LEAMVAAL-DKLKSTITT-----TSPFLLEALEKINTFLGP 1656
QY 123 RRA--VAVRNYLLG 134
DB 1657 HAATLTAITEYCCG 1670
```

```
RESULT 12
US-09-742-659-6
; Sequence 6, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2865
; TYPE: PRT
; ORGANISM: GBV-B/HCV
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2275)..(2865)
; OTHER INFORMATION: chimeric region
; US-09-742-659-6

Query Match 10.2%; Score 87.5; DB 10; Length 2865;
Best Local Similarity 28.4%; Pred. No. 8.8;
Matches 38; Conservative 17; Mismatches 48; Indels 31; Gaps 6;

QY 3 LHIOIAAAAALSVLPMTGCAKSTQVWVAPNAPTGYTGVTGVAPLVNDDETVKAL 62
DB 1566 LAVGVGVAAMAYLAIDFGATCVRKWSITSV---PTGAT-----VAPVDEEIVEEC 1615
QY 63 ASKLPSLVYDFDSDDEIKPQAAAILDEQAQFLTTNOTARVLVAGHTDGRSGREYNMSLGE 122
```

Db 1616 ASFP-----LEAMVAI-DKISTIT-----ISPFLEFALKLNFPUGP 1656
QY 123 RRA--VAVRNYLIG 134
Db 1657 HAATILAIIEYCCG 1670

RESULT 13

US-09-886-468-21
; Sequence 21, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886,468
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 21
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-886-468-21

Query Match 9.5%; Score 82; DB 10; Length 871;
Best Local Similarity 24.0%; Pred. No. 6.2;

Matches 42; Conservative 18; Mismatches 81; Indels 34; Gaps 7;

QY 15 SVLTFF-----TGCAKSTSOVWVAPNAPGTGYTGVAPLVNDDEYKALASKL 66
Db 426 AILAFIDSGSVSDKTGTSTANNNOEVLSTNAATVSGAIVATKCTLTGNS----- 476
QY 67 PSLVYFDPDSDEIKPQAAAILDEQAQFLTTNQTARVYLVAGHTDERSGREY--NMSL-GE 122
Db 477 -----LTFDGNFTAGTSGAIVTETEDFTLTGTVFTSTAKTGALYSKGNNSLSGN 531
QY 123 RRAVAVRNYLLGKGINASVE-----IISGEERPIAFGTNEAM--SQNRREL 171
Db 532 TNLFSGNKATGPNSSANOGCGGAILSPLESASVS--TKGLWIEDENVSLS 584

RESULT 14

US-09-841-132-177
; Sequence 177, Application US/09841132
; Patent No. US20020061848A1

; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FASTSEQ for Windows Version 3.0/4.0
; SEQ ID NO: 177
; LENGTH: 964
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-177

Query Match 9.4%; Score 81; DB 10; Length 964;
Best Local Similarity 23.6%; Pred. No. 9.2;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLFPMTCANKSTSOVWVAPNAPGTGYTGVAPLVNDDEYKALASK 65
Db 277 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNTGGAIVAVTVLVNCGPTFIINNIA 334
QY 66 LPSLVYFDPDSDEIKPQAAAILDEQAQFLTTNQTARVYLVAGHTDERSREYNMSLGERRA 125
Db 335 KGGAIYIDGTSN-----SKISADRHAIIFENNIYNTVNNAGTSTANP-----RNA 383
QY 126 VAVRN-----YLLGKGINO-----ASVELISGEERPIAFGTNEA 161
Db 384 ITVASSSGEILLGAGSSQNLIFYPDIEVSNAG---VVSFNKEA 424

RESULT 15

US-09-841-132-191
; Sequence 191, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FASTSEQ for Windows Version 3.0/4.0
; SEQ ID NO 191
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-191

Query Match 9.4%; Score 81; DB 10; Length 977;
Best Local Similarity 23.6%; Pred. No. 9.4;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLFPMTCANKSTSOVWVAPNAPGTGYTGVAPLVNDDEYKALASK 65
Db 290 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNTGGAIVAVTVLVNCGPTFIINNIA 347
QY 66 LPSLVYFDPDSDEIKPQAAAILDEQAQFLTTNQTARVYLVAGHTDERSREYNMSLGERRA 125
Db 348 KGGAIYIDGTSN-----SKISADRHAIIFENNIYNTVNNAGTSTANP-----RNA 396
QY 126 VAVRN-----YLLGKGINO-----ASVELISGEERPIAFGTNEA 161
Db 397 ITVASSSGEILLGAGSSQNLIFYPDIEVSNAG---VVSFNKEA 437

RESULT 16

US-10-007-693-74
; Sequence 74, Application US/10007693
; Patent No. US2002014676A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 74
; LENGTH: 361
; TYPE: PRF
; ORGANISM: Chlamydia pneumoniae
US-10-007-693-74

Query Match
Best Local Similarity 24.5%; Score 80; DB 12; Length 361;
Matches 37; Conservative 27; Mismatches 43; Indels 44; Gaps 8;

QY 12 AALSVLTFTGTCANKSTQVWAPNA-PTG-----YTGVTYGVAPLVNDETVALAKSL 66
DB 158 SALHNVLPPLGEMNTEVRAIAQAALPTAEKSTGICIGKRFKFELE--KFLPKKT 215
QY 67 PSLYVEFDESDDEIKFOAAIIDEQAQFLTTQTARVLVAAGHIDERSREYMSIGERRAV 126
DB 216 GNVV--DWDTRKIVQ-----HOGAHY--TIGORRGL 244

QY 127 AVR-----YLLGKINGASVEIISFGGERP 152
DB 245 DLGSEKPCYVVGKNIENSIYIVR-GEDHP 274

RESULT 17
US-10-156-761-9543
; Sequence 9543, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9543
; LENGTH: 420
; TYPE: PRF
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9543

Query Match
Best Local Similarity 23.4%; Score 80; DB 9; Length 420;
Matches 43; Conservative 33; Mismatches 84; Indels 24; Gaps 6;

QY 6 QIAAAALSVLTFTGTCANKSTQVWAPNAPGTYGVITGVAPLVNDETVALAKSL 62
DB 172 EIAAGVGLTIDTLGLVNSKLSKLSIAEA---TSGYTSVHKEDLDKYNQGLVDR 227
QY 63 -ASKLPSLYVEFDESDDEIKFOA--AAIIDEQAQF-----LTTNQTARVLVAAGHIDE 110
DB 228 AADKVVTVAVDGAADCAKAPLPLASGLFTTDRAEFAQHRWYVDVAPGDELASVSLADR 287

QY 111 RGSREYNMSIGERRAVARN--YLLGKINGASVEIISGERRIAPGTNEAMSONRRA 168
DB 288 QVNPDIQVLL---RAVTHNRREIYRGEMAGTGTVDVSTGLRYPKASDEQDATAETVCL 344

QY 169 ELAY 172
DB 345 QVSH 348

RESULT 18
US-10-156-761-13758
; Sequence 13758, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13758
; LENGTH: 877
; TYPE: PRF
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13758

Query Match
Best Local Similarity 24.1%; Score 77.5; DB 9; Length 877;
Matches 53; Conservative 23; Mismatches 87; Indels 57; Gaps 8;

QY 10 AAAASVLTFTGTCANKSTQVWAPNAPT--GYTGVITYGVA--PLVNDDETVALAS- 64
DB 268 APAAGVDAVLPAGGRVVLATSVASSSLTPGVRRVVOGLAREPRVDHARLSLITV 327
QY 65 -----KLPSLYV-----DPSDEIKP-----QAAIIDE 89
DB 328 RASQAGRORAGRAGREAPGTVYRCWAEBADARLPFRPSPETIKVADLTAFAAOACWGP 387
QY 90 QAGFLT-----TNTQARVLV-----AGHTDERSREYNMSIGERRAVARNRYLL 133
DB 388 EASGLALLDPPGGAANAASVLAAGANDSAGRATERGVRSRLGHPRLARALLDAAP 447

QY 134 GKGINGASVEIISFGERRIAPGTN--EAMSONRRAELSY 172
DB 448 EVGADRAAEVVALISEPPEYGGDDLAAARARRGDAV 487

RESULT 19
US-09-815-242-14076
; Sequence 14076, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes

CLASSIFICATION: <Unknown>

137 AYPDWYI TQVCTNOACVET

—
—
..
—

..
..
..
—
..
..
..
..
..

—
—
—
—

Db 98 G-GFYLWG-----AKWVAVGDQKVVKTLRGQLGTMEEGTS 132

RESULT 22

US-09-815-242-5697

Sequence 5697, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5697

LENGTH: 794

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-09-815-242-5697

Query Match 8.8%; Score 75.5; DB 10; Length 794;

Best Local Similarity 23.8%; Pred. No. 27;

Matches 38; Conservative 24; Mismatches 59; Indels 39; Gaps 6;

QY 9 AAAALSVLFMTGCA--NKSTSQVWVAPN--APTGY-----TG 43

Db 439 ALVASISVLYIACPCALGLATPTSIMVGTGAENGILFKGGEVERTHQIDTIVLDKGTG 498

QY 44 VIYGVAPLVD---NDEYKALASKLPISLYFEDSDSEIKRQAAALID--EQAQFLTTNQ 98

Db 499 TITNGRPVYTDYHGDQDTQLALATA-----EKDSEHPLAEALIVNAKKEKQLTLTET 549

QY 99 TARVLVAGHTDEGRSREYNMSLGERRAVAVRNLYLGKGIN 138

Db 550 TTFKAVPGHGIEATIDHHILVGNKRLMADNDISLPKHIS 589

RESULT 23

US-09-815-242-12668

Sequence 12668, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 12668

LENGTH: 802

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-09-815-242-12668

Query Match 8.8%; Score 75.5; DB 10; Length 802;

Best Local Similarity 23.8%; Pred. No. 27;

Matches 38; Conservative 24; Mismatches 59; Indels 39; Gaps 6;

QY 9 AAAALSVLFMTGCA--NKSTSQVWVAPN--APTGY-----TG 43

Db 439 ALVASISVLYIACPCALGLATPTSIMVGTGAENGILFKGGEVERTHQIDTIVLDKGTG 498

QY 44 VIYGVAPLVD---NDEYKALASKLPISLYFEDSDSEIKRQAAALID--EQAQFLTTNQ 98

Db 499 TITNGRPVYTDYHGDQDTQLALATA-----EKDSEHPLAEALIVNAKKEKQLTLTET 549

QY 99 TARVLVAGHTDEGRSREYNMSLGERRAVAVRNLYLGKGIN 138

Db 550 TTFKAVPGHGIEATIDHHILVGNKRLMADNDISLPKHIS 589

RESULT 24

US-10-156-761-10510

Sequence 10510, Application US/10156761

Patent No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

PRIOR FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272667

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 10510

LENGTH: 1330

TYPE: PRT

ORGANISM: Streptomyces avermitilis

US-10-156-761-10510

Query Match

Best Local Similarity 8.7%; Score 74.5; DB 9; Length 1330;

Matches 25.0%; Pred. No. 72;

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4444
LENGTH: 1520
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4444

Query Match
Best Local Similarity 24.6%; Pred. No. 1.3e+02;
Matches 28; Conservative 25; Mismatches 47; Indels 14; Gaps 5;

QY 53 VDNDETVKALASKLPSLYVYFDPDSDEIKPOAAIIDEQAFLLTNOTA----RVLVAGH- 107
DB 311 VNSRRSAERLTSRLNEIWMHEHDPESLSPQ---LRDPQAIMSSADVAGKAPQVYARAH 367
QY 108 ----TDERGSRREYNMSLGERRAAVAVRNLLGKINGASVEIISFGGERP-TAFG 156
DB 368 GSVSKDERATETMTLKEGRRLAV-ISTSLLEIGIDMGAVDLVYOVESPVSASG 420

RESULT 33

US-10-156-761-9933
Sequence 9933, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9933
LENGTH: 851
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9933

Query Match
Best Local Similarity 21.1%; Pred. No. 62;
Matches 45; Conservative 19; Mismatches 60; Indels 89; Gaps 8;

QY 6 QIAAAALSLVTFEMGCKAKSTSOV-----MVPNATGYT 42
DB 483 QAAALRAESELISFLAGSVLRGETSLDALLERYETFGMESVALLERESDVAP----WT 537
QY 43 GVIYTVGAVPLVD-----NDETVALASK-LPSLYVFPDSDE----IKPOAA 85
DB 538 CAGSVGYRPLVDREDAVDMPGDHALLASGRVLP-----SDRRYLAFAQAANV 590
QY 86 ILDEQAFLLTNOTARVLVAGHTDERGSRREYNMSLGERRAAVAVRNLLGKINGASVEI 145
DB 591 VLDRORL-----QREDAQRTLAEGNSITALLAAV 621
QY 146 SFGGERPIA-----FGTNEAMSONRRRAEL 170
DB 622 SHDLRTPLAGIKAAVSSLRSDVAVMSSEDRAL 654

RESULT 34

US-09-901-106-2
Sequence 2, Application US/09901106
Patent No. US20020151067A1
GENERAL INFORMATION:

APPLICANT: Garoff, Henrik
Liljestrom, Peter
TITLE OF INVENTION: DNA Expression Systems Based on
Alphaviruses
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/901,106
FILING DATE: 10-Jul-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/920,281C
FILING DATE: 13-AUG-1992

ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 2431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-901-106-2

Query Match
Best Local Similarity 23.0%; Pred. No. 2.8e+02;
Matches 26; Conservative 20; Mismatches 36; Indels 31; Gaps 5;

QY 24 ANKSTQYVAVAPNMPGTGYITGVAVPLVNDETVALASKLPSLYVFPDSDEIKPOA 83
DB 1756 ASRAERVPAPARKPT-----PAKRT-----AFRNKLP-LTFGDFDEHYVALA 1798
QY 84 AAI-----LDEQAFLLTNOTARVLVAGHTDERGSRREYNMSLGERRAV 126
DB 1799 SGITFGDFPDVLRGRAGAVYIFSSDTG-----SGHLQKQSVQHNLCQAQLDAAV 1847

RESULT 35

US-10-156-761-13680
Sequence 13680, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13680
LENGTH: 441
TYPE: PRT
ORGANISM: Streptomyces avermiltillis
US-10-156-761-13680

Query Match 8.3%; Score 71.5; DB 9; Length 441;
Best Local Similarity 24.1%; Pred. No. 31;
Matches 46; Conservative 21; Mismatches 73; Indels 51; Gaps 7;

QY 6 QIAAAALSVLTFTMTGKANKSTQYVAPN-----APTGYTVI-----YTGVAL 52
DB 186 EYVAGGATVAVAFVGTSCAPADVTGPGNIVAAAKRYFTGKIGIDAEAGPTETIATL 245
QY 53 VDNDEVKALASL-----PSLVTFDESDKLPQAAA---ILDEQAF 93
DB 246 ADSTADPVHVASDLISQAEHDLAAAVLYTDSVELDAVEKELEPQVATKHIIDRIYPA 305
QY 94 LTTNOTARVLVAG-----HTDERSREYNMSLGERRAVA--VRNYLGGKINQASVEII 145
DB 306 LKRGSAIYLVYDGVDSGLRVADYAGAEHLEIQTADAAVADRVRN-----AGAI 354
QY 146 SFGEERPIAFG 156
DB 355 FIGMPAPVSLG 365

RESULT 36
US-10-156-761-14476
Sequence 14476, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14476
LENGTH: 507
TYPE: PRT
ORGANISM: Streptomyces avermiltillis
US-10-156-761-14476

Query Match 8.3%; Score 71.5; DB 9; Length 507;
Best Local Similarity 23.5%; Pred. No. 38;
Matches 32; Conservative 25; Mismatches 50; Indels 29; Gaps 6;

QY 41 YTGVIITGVAPLVNDDEVKALASL-----PSLVTFDESD-----EIKP 81
DB 245 FTGGLTGRIMAAAGTAKVLELGGKNPNIVFADADFDPAVDMALAVFLHSQVCS 304
QY 82 QAAAILDEQA---QFL--TTNCTARVLVAGHIDERSREYNMSLGERRAVAVRNYLGGK 136
DB 305 AGARLLVEDSLHDFVDEVVRAAEIRLGGPDERAQTGALISAARAVEA---YVANG 361
QY 137 INQASVEIISFGEERP 152
DB 362 LDEGAV--LRGGRRP 375

RESULT 37
US-09-742-659-5
Sequence 5, Application US/09742659
Patent No. US20010034019A1
GENERAL INFORMATION:
APPLICANT: Hong, Zhi
APPLICANT: Butkiewicz, Nancy J.
APPLICANT: Zhong, Weidong
APPLICANT: Ingravallo, Paul
APPLICANT: Wright-Minogue, Jacquelyn
APPLICANT: Lau, Johnson Y.
APPLICANT: Lemon, Stanley M.
TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
FILE REFERENCE: ID01116
CURRENT APPLICATION NUMBER: US/09/742,659
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: US 60/171,469
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 2862
TYPE: PRT
ORGANISM: GBV-B/HCV
FEATURE:
NAME/KEY: SITE
LOCATION: (945)..(1129)
OTHER INFORMATION: chimeric region
NAME/KEY: SITE
LOCATION: (1579)..(1593)
OTHER INFORMATION: chimeric region
US-09-742-659-5

Query Match 8.3%; Score 71.5; DB 10; Length 2862;
Best Local Similarity 27.5%; Pred. No. 4.5e+02;
Matches 38; Conservative 16; Mismatches 43; Indels 41; Gaps 8;

QY 3 LHIQIAAALSVLTFTMTGKANKSTQYVAPN-----PAPTYGYTVITGVAPLVNDDET 58
DB 1566 LAVGVGAMATLA-----STGCV-----VYGRVYLSGKPLGAT-----VAPVDEEII 1609
QY 59 VKALASKLPSLVTFDESDKLPQAAAILDEQAQFLTTNCTARVLVAGHIDERSREYNM 118
DB 1610 VEECASFP-----DEAMVAAT-DLKSTITT-----TSPFTLEFALERTLT 1650
QY 119 SLGERRA--VAVRNYLGG 134
DB 1651 FLGPHRAATLAIIEYCG 1668

RESULT 38
US-09-925-301-908
Sequence 908, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 908
LENGTH: 248
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-301-908

Query Match 8.2%; Score 71; DB 10; Length 248;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:16:24 ; Search time 62 Seconds
(without alignments)
850.781 Million cell updates/sec

Title: US-09-674-779B-2
Perfect score: 861
Sequence: 1 MMLHIGIAAAALSVLTFM.....IAFGTNEAASQNRRAELSY 172

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Command line parameters:
-MODEL=frame+P2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09674779/rnalt.06072003.121614.9825/app_query.fasta.1.327
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=100 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTEXT=plto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09674779.ecgn.1.1.36.gunat -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -NCPU=6 -ICPU=3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/1/lna/6C.COMB.seq:*
6: /cgn2_6/prodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	30.7	462	2	US-08-743-637B-178
2	264	30.7	462	6	5173294-3
3	264	30.7	867	6	5173294-1
4	264	30.7	1019	1	US-07-807-049-1
5	176.5	20.5	1059	4	US-08-476-102A-3
6	175.5	20.4	645	2	US-08-572-447C-12
7	175.5	20.4	645	4	US-09-267-747-12
8	171.5	19.9	486	2	US-08-572-447C-10
9	171.5	19.9	486	4	US-09-267-747-10
10	171.5	19.9	681	2	US-08-572-447C-14
11	171.5	19.9	681	4	US-09-267-747-14
12	165.5	19.2	1720	1	US-08-457-997B-1

13	165.5	19.2	1720	3	US-08-467-722A-1	Sequence 1, Appl1
14	144	16.7	3656	4	US-09-221-017B-786	Sequence 786, App
15	143	16.6	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
16	143	16.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
17	142	16.5	4760	4	US-09-103-840B-1	Sequence 875, App
18	136.5	15.9	730	2	US-08-743-637B-11	Sequence 11, Appl
19	136.5	15.9	730	2	US-08-526-840B-11	Sequence 11, Appl
20	136.5	15.9	1007	4	US-08-836-500A-1	Sequence 13, Appl
21	136.5	15.9	1008	3	US-08-721-979A-13	Sequence 13, Appl
22	136.5	15.9	1008	4	US-08-654-289-13	Sequence 13, Appl
23	136.5	15.8	1727	1	US-08-129-719-14	Sequence 14, Appl
24	136	15.8	1727	1	US-08-306-871-14	Sequence 14, Appl
25	136	15.8	1727	1	US-08-569-959-14	Sequence 14, Appl
26	87.5	10.2	4268	4	US-08-469-260A-80	Sequence 80, Appl
27	87.5	10.2	8912	2	US-08-469-260A-390	Sequence 11, Appl
28	87.5	10.2	9143	2	US-08-469-260A-393	Sequence 32, Appl
29	87.5	10.2	9143	4	US-08-469-260A-393	Sequence 390, App
30	87.5	10.2	9143	4	US-08-469-260A-393	Sequence 393, App
31	81	9.4	2895	4	US-09-556-877-171	Sequence 171, App
32	81	9.4	2895	4	US-09-556-877-171	Sequence 171, App
33	81	9.4	2934	4	US-09-620-412C-183	Sequence 183, App
34	81	9.4	2934	4	US-09-620-412C-183	Sequence 183, App
35	78	9.1	11958	4	US-09-134-246-8	Sequence 13, Appl
36	77.5	9.0	80161	3	US-09-036-987A-1	Sequence 17, App
37	77.5	9.0	80161	3	US-09-036-987A-1	Sequence 17, App
38	77	8.9	4403765	4	US-09-103-840A-2	Sequence 183, App
39	74	8.6	38155	4	US-09-453-702B-137	Sequence 137, App
40	74	8.6	48908	4	US-09-453-702B-137	Sequence 137, App
41	73.5	8.5	1242	2	US-08-762-106-6	Sequence 6, Appl1
42	73.5	8.5	1581	1	US-09-320-774-6	Sequence 6, Appl1
43	73.5	8.5	1581	1	US-09-320-774-6	Sequence 6, Appl1
44	73.5	8.5	5849	4	US-09-134-246-6	Sequence 44, Appl
45	73.5	8.5	8976	4	US-09-495-797-44	Sequence 4, Appl1
46	73	8.5	6201	2	US-08-790-912-1	Sequence 1, Appl1
47	73	8.5	6201	2	US-08-790-912-1	Sequence 1, Appl1
48	73	8.5	12848	4	US-09-453-702B-152	Sequence 25, App
49	72.5	8.4	11517	1	US-07-920-281C-1	Sequence 1, Appl1
50	72.5	8.4	11517	1	US-08-466-277-1	Sequence 1, Appl1
51	72	8.3	12492	6	5206163-2	Sequence 1, Appl1
52	71.5	8.3	1108	1	US-08-181-271A-13	Sequence 13, Appl
53	71.5	8.3	1108	1	US-08-449-315-13	Sequence 13, Appl
54	71.5	8.3	1108	1	US-08-449-315-13	Sequence 13, Appl
55	71.5	8.3	1108	1	US-08-449-315-13	Sequence 13, Appl
56	71.5	8.3	1108	1	US-08-449-315-13	Sequence 13, Appl
57	71.5	8.3	1108	1	US-08-449-315-13	Sequence 13, Appl
58	71.5	8.3	1108	1	US-08-449-315-13	Sequence 13, Appl
59	71.5	8.3	1108	1	US-08-449-315-13	Sequence 13, Appl
60	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
61	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
62	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
63	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
64	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
65	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
66	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
67	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
68	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
69	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
70	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
71	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
72	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
73	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
74	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
75	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
76	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
77	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
78	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
79	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
80	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
81	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
82	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
83	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
84	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, Appl
85	69	8.0	2797	2	US-08-954-333-8	Sequence 8, Appl1

86 69 8.0 2820 5 PCT-US94-05905-19 Sequence 19, Appl
87 69 8.0 5117 3 US-08-854-585-1 Sequence 1, Appl
88 69 8.0 5117 5 PCT-US95-05512-1 Sequence 1, Appl
89 69 8.0 531 2 US-08-809-267-9 Sequence 9, Appl
90 68.5 8.0 531 5 PCT-US95-13662A-9 Sequence 9, Appl
91 68.5 8.0 1359 1 US-07-618-312A-3 Sequence 7, Appl
92 68.5 8.0 1359 1 US-08-110-786A-7 Sequence 6, Appl
93 68.5 8.0 1359 1 US-08-280-228-3 Sequence 39, Appl
94 68.5 8.0 1858 1 US-08-668-381A-6 Sequence 57, Appl
95 68.5 8.0 25165 4 US-09-453-702B-39 Sequence 60, Appl
96 68 7.9 955 3 US-08-784-582-60 Sequence 72, Appl
97 68 7.9 2356 3 US-08-784-582-72 Sequence 314, Appl
98 68 7.9 3190 4 US-09-221-017B-314 Sequence 242, App
99 68 7.9 31880 4 US-09-453-702B-242

ALIGNMENTS

RESULT 1
US-08-743-637B-178
Sequence 178, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESSES:
ADDRESSEE: OUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5500
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 178:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Haemophilus influenzae
US-08-743-637B-178
Alignment Scores: 9,21e-27 Length: 462
Pred. No.: 264.00 Matches: 51
Score: 264.00

Percent Similarity: 68.93%
Best Local Similarity: 49.51%
Query Match: 30.66%
DB: 2 Gaps: 0

US-09-674-779B-2 (1-172) x US-08-743-637B-178 (1-462)

70 ValTyrPheAspPheAspSerAspGluIleTyrProGlnAlaAlaIleLeuAspGlu 89
151 GRTATATTGGTTTGATTAATACGACATCCGCGTGAATACGTTCAATCTTAGATCG 210
90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaValAlaGlyHisThrAsp 109
211 CACGACGATATTAAATGCAACGCCAGCTGCTAAAGTATGTTGAGAGTAACTGAT 270
110 GlnArgGlySerArgGluTyrThrAsnMetSerLeuGlyGlnArgAlaValAlaValArg 129
271 GAACGTGTACACCAATACCAATACCAATACCAATACCAATACCAATACCAATACCAAT 330
130 AsnTyrLeuLeuGlyGlyLeuGlnAlaSerValGluIleLeuSerPheGlyGlu 149
331 GGTATTATTACGAGTAAGGTGTGATGCTGTAATTAAGGACACGATATCTTACGGTGA 390
150 GlnArgProIleAlaPheGlyThrAsnGlnGluAlaIlePheSerGlnAsnArgAlaGlu 169
391 GAAACACCTGACATATTAGTGTACGATGAGAGCTGCAATCTTAAACCGTGTGACAGT 450
170 LeuSerTyr 172
451 TTACGCTAC 459

RESULT 2

5173294-3
Patent No. 5173294
APPLICANT: MURPHY, TIMOTHY F.; APICELLA, MICHAEL A.
TITLE OF INVENTION: DNA PROBE FOR THE IDENTIFICATION
OF HAEMOPHILUS INFLUENZAE
NUMBER OF SEQUENCES: 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/92,948
FILING DATE: 08-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 932,872
FILING DATE: 18-NOV-1986
SEQ ID NO: 3
LENGTH: 462
5173294-3

Alignment Scores:

Pred. No.: 9,21e-27 Length: 462
Score: 264.00 Matches: 51
Percent Similarity: 68.93% Conserved: 20
Best Local Similarity: 49.51% Mismatches: 32
Query Match: 30.66% Indels: 0
DB: 6 Gaps: 0

US-09-674-779B-2 (1-172) x 5173294-3 (1-462)

70 ValTyrPheAspPheAspSerAspGluIleTyrProGlnAlaAlaIleLeuAspGlu 89
151 GRTATATTGGTTTGATTAATACGACATCCGCGTGAATACGTTCAATCTTAGATCG 210
90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaValAlaGlyHisThrAsp 109
211 CACGACGATATTAAATGCAACGCCAGCTGCTAAAGTATGTTGAGAGTAACTGAT 270
110 GlnArgGlySerArgGluTyrThrAsnMetSerLeuGlyGlnArgAlaValAlaValArg 129
271 GAACGTGTACACCAATACCAATACCAATACCAATACCAATACCAATACCAATACCAAT 330
130 AsnTyrLeuLeuGlyGlyLeuGlnAlaSerValGluIleLeuSerPheGlyGlu 149
331 GGTATTATTACGAGTAAGGTGTGATGCTGTAATTAAGGACACGATATCTTACGGTGA 390

QY 150 GUARGProlleAlaPheGlyThrAsnGluLualatrpSerGlnAsnArgAlaGlu 169
 |||:|||||
 Db 391 GAAAAACCTGCAGATTAGTACGATGAAAGCTGCATATTCTAAACCGTCTGCAGTG 450
 |||:|||||
 QY 170 LeuSerTyr 172
 |||:|||||
 Db 451 TTAGCGTAC 459
 |||:|||||

RESULT 3
 5173294-1
 Patent No. 5173294
 APPLICANT: MURPHY, TIMOTHY F., APICELLA, MICHAEL A.
 TITLE OF INVENTION: DNA PROBE FOR THE IDENTIFICATION
 OF HAEMPHILUS INFLUENZAE
 NUMBER OF SEQUENCES: 3
 CURRENT APPLICATION DATA:
 FILING DATE: 08-OCT-1987
 APPLICATION NUMBER: US/07/92,948
 PRIOR APPLICATION DATA:
 FILING DATE: 18-NOV-1986
 APPLICATION NUMBER: 932,872
 SEQ ID NO.1:
 LENGTH: 867
 5173294-1

Alignment Scores:
 Pred. No.: 2.38e-26 Length: 867
 Score: 264.00 Matches: 51
 Percent Similarity: 68.93% Conservative: 20
 Best Local Similarity: 49.51% Mismatches: 32
 Query Match: 30.66% Indels: 0
 DB: 6 Gaps: 0

US-09-674-779b-2 (1-172) x 5173294-1 (1-867)

QY 70 ValTyrPheAspPheSerAspGluIleYsProGlnAlaAlaIleLeuAspGlu 89
 |||:|||||
 Db 218 GTATATTGCTTTTATTAATACGATCCGCGTGAATACGTTCAATCTTGATGCG 277
 |||:|||||

QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
 |||:|||||
 Db 278 CACGCGACGATATTTAATGCAACGCCAGCTGCTAAGATTAGTGAAGGAAATACGAT 337
 |||:|||||

QY 110 GluArgGlySerArgIleuTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 |||:|||||
 Db 338 GAACGCTGATACACCGAATACACATCCGATTAGGACAAACGTCGTCAGATGCAATTAA 397
 |||:|||||

QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
 |||:|||||
 Db 398 GGTATTATTAGCAGGTAAGGCTGTGATGCTGTAATATTAGGCACAGATCTTACGGTGA 457
 |||:|||||

QY 150 GUARGProlleAlaPheGlyThrAsnGluLualatrpSerGlnAsnArgAlaGlu 169
 |||:|||||
 Db 458 GAAAAACCTGCAGATTAGTACGATGAAAGCTGCATATTCTAAACCGTCTGCAGTG 517
 |||:|||||

QY 170 LeuSerTyr 172
 |||:|||||
 Db 518 TTAGCGTAC 526
 |||:|||||

RESULT 4
 US-07-807-049-1
 Sequence 1, Application US/07807049
 Patent No. 5300632
 GENERAL INFORMATION:
 APPLICANT: Murphy, Timothy F
 APPLICANT: Apicella, Michael A
 TITLE OF INVENTION: A Method for Purifying an Outer Membrane
 PROTEIN OF HAEMOPHILUS INFLUENZAE
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Alan S. Korman, Esq.
 STREET: 1600 Empire Tower

CITY: Buffalo
 STATE: NY
 COUNTRY: U.S.A.
 ZIP: 14202

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/807,049
 FILING DATE: 19911212
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/330,229
 FILING DATE: 29-MAR-1989

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/092,948
 FILING DATE: 08-OCT-1987

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/932,872
 FILING DATE: 18-NOV-1986

ATTORNEY/AGENT INFORMATION:
 NAME: KORMAN ESQ., ALAN S
 REGISTRATION NUMBER: 32,932

REFERENCE/DOCKET NUMBER: 19226/00028

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 853-8100
 TELEFAX: (716) 853-8109

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1019 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO

ORGANISM: Haemophilus influenzae
 STRAIN: 1479
 INDIVIDUAL ISOLATE: 1479

HAPLOTYPE: N/A
 IMMEDIATE SOURCE:
 LIBRARY: N/A

CLONE: N/A
 POSITION IN GENOME:
 MAP POSITION: Unknown

PUBLICATION INFORMATION:
 AUTHORS: Nelson, M B
 AUTHORS: Apicella, M A
 AUTHORS: Murphy, T F
 AUTHORS: Vankulien, H
 AUTHORS: Spetella, L D
 AUTHORS: Rekosh, D M

TITLE: Molecular Analysis of P6: The cloning and
 TITLE: sequencing of an Outer Membrane Protein
 JOURNAL: Infect. Immun.
 VOLUME: 56
 PAGES: 128-134

DATE: 1988
 PUBLICATION INFORMATION:
 DOCUMENT NUMBER: US 4,427,782
 FILING DATE: 24-JAN-1984

PUBLICATION INFORMATION:
 DOCUMENT NUMBER: US 4,474,758
 FILING DATE: 02-OCT-1984

PUBLICATION INFORMATION:
 DOCUMENT NUMBER: US J.CLIN.INVEST.V.78
 PUBLICATION DATE: 01-OCT-1986

RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1020 TO 1027


```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/572.447C
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 645 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: linear
HYPOTHETICAL: NO
ANTI-SENSE: YES
ORGANISM: Pseudomonas aeruginosa
TISSUE TYPE: Serotype 6; ATCC 33354
FEATURE:
NAME/KEY: CDS
LOCATION: 1..642
OTHER INFORMATION: /note= "Sequence is coding for oprf
; Patent No. 5955090
; OTHER INFORMATION: C-terminus and oprf without signal sequence"
US-08-572-447C-12

Alignment Scores:
Pred. No.: 1,41e-14 Length: 645
Score: 175.50 Matches: 49
Percent Similarity: 46.58% Conservative: 19
Best Local Similarity: 33.56% Mismatches: 68
Query Match: 20.38% Gaps: 11
DB: 2 Gaps: 2

US-09-674-779b-2 (1-172) x US-08-572-447C-12 (1-645)
QY 37 AlaProthGlyTyrThrGlyValAlaProLeuValaspasasp 56
    |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 26 GCTCCGACTCCGACACAGCGGCTCTGCGACAC-GTCGACAAAGTCCCGACACCCCG 84
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 57 GluThrValLysAlaLeuAlaSerLysLeuProSerLeu----- 69
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 85 GCCAAGCTACCGCTTACCGCAACGGCTCCCGGCTGCGCGAAGTCTACCGCTACAG 144
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 70 -----ValTyrPheaspPheaspSerAspGluLeuProGluAlaAlaLeu 87
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 145 CTGAGAGTGAAGTTCGACTTCGACAAAGTCCAAAGTCAAGACAGCTACCTGACATC 204
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 88 AspGluAlaGlnPheLeuThrThrAsnGlnThrAlaAspValLeuValaGlyHis 107
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 205 AAGAACCTGGCGACTTCATGAGAGCAAGTACCCGCTCCTCCACCAACGTTGAAGTCA 264
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 108 ThrAspGluArgGlySerArgLutTyrAsnMetSerLeuGlyLysArgAlaValAla 127
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 265 ACCGACTCGGTGAGTACGACGCTTACCAACGAAAGCTGTCCGAGCGTGTCACCAAGCC 324
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 128 ValArgAsnTyrLeuLeuGlyLys---GlyTyrLeuGlnAlaSerValGluLeuSer 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 325 GTTCGAGCTACTGTGTCAACGATGATGGAAGTGTCGCGTGAAGCTGTGCGT 384
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 147 PheGlyGluGluArgProLeuAlaPheGlyThrAsnGluGluAlaTyrSerGlnAsnArg 166
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```

DB 385 TAGCGAGATCCCGCGGTGTCGACAAACGCCACCGCTGAAGCCGCGCTATCAACCGT 444
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 167 ArgAlaGluLeuSerTyr 172
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 445 CGCGTTGMAAGCAGCCAC 462

RESULT 7
US-09-267-747-12
; Sequence 12, Application US/09267747
; Patent No. 6300102
GENERAL INFORMATION:
APPLICANT: Kaapp, Bernhard
APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Broker, Michael
APPLICANT: Von Specht, Bernd-Ulrich
APPLICANT: Domdey, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-Oprf
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/267,747
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/572,447
FILING DATE: 14-DEC-1995
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 645 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: linear
HYPOTHETICAL: NO
ANTI-SENSE: YES
ORGANISM: Pseudomonas aeruginosa
TISSUE TYPE: Serotype 6; ATCC 33354
FEATURE:
NAME/KEY: CDS
LOCATION: 1..642
OTHER INFORMATION: /note= "Sequence is coding for oprf
; Patent No. 6300102
; OTHER INFORMATION: C-terminus and oprf without signal sequence"
US-09-267-747-12

Alignment Scores:
Pred. No.: 1,41e-14 Length: 645
Score: 175.50 Matches: 49
Percent Similarity: 46.58% Conservative: 19
```

Best Local Similarity: 33.56% Mismatches: 68
 Query Match: 20.38% Indels: 11
 DB: 4 Gaps: 2

US-09-674-779b-2 (1-172) x US-09-267-747-12 (1-645)

QY 37 AlaProthrglytyrthrglyvallelyrthrglyalaProleuValaAspaansp 56
 |||||
 DB 26 GCTCCGACTCCGACAGAGGCGCTCTGCGACAC-GTCGACAAAGTCCCGGACACCCG 84
 QY 57 GluThrVallyslalaLeuAlaSerlyslLeuProserLeu----- 69
 |||||
 DB 85 GCCACGTGACCGCTTACCGCCCAACGCGCTCCGCGCTCCGGAAGTCGTACGCGTACAG 144
 QY 70 -----ValTyrPheaspPheaspseraspGluilelyProglAlaAlaIleleu 87
 |||||
 DB 145 CTGGAGCTGAGTACTTGCACACTTGCACAGTCCAAAGTCAAGAACAGCTACGCTGACATC 204
 QY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValleuValaAglyHis 107
 |||||
 DB 205 AACAGCTGCGCGACTTCATGAAAGCAGTACCGCTCCACTTCACACCGCTTGAAGTCAAT 264
 QY 108 ThrAspGluArglyserArglyTyrAsnMetSerLeuGlyGluArgAlaValaAla 127
 |||||
 DB 265 ACCGACTCCGCTCGTACCGCAGCTTACCAACGACGAGCTTCCGAGGCTGCTCCACAGCC 324
 QY 128 ValArgAspTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleIleSer 146
 |||||
 DB 325 GTTCGTGACGTACTGTGCAACGAGTACGAGTGTGAAGTGTGCGCTGAACCTGTGCGGT 384
 QY 147 PheGlyGluArgProIleAlaPheGlyThrAsnGluGlnAlaTrpSerGlnAsnArg 166
 |||||
 DB 385 TACGGGAGTCCGCGCGCTGCGCGACCAACGCGCAGGCGCGCTATCAACCGT 444
 QY 167 ArgAlaGluLeuSerTyr 172
 |||||
 DB 445 CGCGTTGAAGCAGCCAC 462

RESULT 8

US-08-572-447C-10
 ; Sequence 10, Application US/08572447C
 ; Patent No. 5955090

GENERAL INFORMATION:
 ; APPLICANT: Knapp, Bernhard
 ; APPLICANT: Hungerer, Klaus-Dieter
 ; APPLICANT: Broker, Michael
 ; APPLICANT: Von Specht, Bernd-Ulrich
 ; APPLICANT: Domdey, Horst
 ; TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-Oprf
 ; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
 ; ADDRESS: Dunnet
 ; STREET: 1300 I Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/572.447C
 ; FILING DATE: 14-DEC-1995
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: EP 94120023.0
 ; FILING DATE: 16-DEC-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: W. Paul Barker

REGISTRATION NUMBER: 32,013
 ; REFERENCE/DOCKET NUMBER: 05552.1395-00000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 408-4000
 ; TELEFAX: (202) 408-4400
 ; INFORMATION FOR SRO ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 486 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Pseudomonas aeruginosa
 ; TISSUE TYPE: Serotype 6; ATCC 33354
 ; FEATURE:

NAME/KEY: CDS
 ; LOCATION: 1..483
 ; OTHER INFORMATION: /note= "Sequence is coding for oprf"
 ; PATENT NO. 5955090
 ; OTHER INFORMATION: C-terminus"
 ; US-08-572-447C-10

Alignment Scores:
 ; Pred. No.: 3,19e-14 Length: 486
 ; Score: 171.50 Matches: 48
 ; Percent Similarity: 46.15% Conservative: 18
 ; Best Local Similarity: 33.57% Mismatches: 67
 ; Query Match: 19.92% Indels: 11
 ; DB: 2 Gaps: 2

US-09-674-779b-2 (1-172) x US-08-572-447C-10 (1-486)

QY 37 AlaProthrglytyrthrglyvallelyrthrglyalaProleuValaAspaansp 56
 |||||
 DB 32 GCTCCGACTCCGACAGAGGCGCTCTGCGACAC-GTCGACAAAGTCCCGGACACCCG 90
 QY 57 GluThrVallyslalaLeuAlaSerlyslLeuProserLeu----- 69
 |||||
 DB 91 GCCAAGCTGACCGTGTGACGCGCACGCGCTGCGCGGAGTCCGAGTCCGCTACAG 150
 QY 70 -----ValTyrPheaspPheaspseraspGluilelyProglAlaAlaIleleu 87
 |||||
 DB 151 CTGGAGCTGAGTACTTGCACACTTGCACAGTCCAAAGTCAAGAACAGCTACGCTGACATC 210
 QY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValleuValaAglyHis 107
 |||||
 DB 211 AAGAACCTGCGCGACTTCATGAAAGCAGTACCGCTCCACTTCACACCGCTTGAAGTCAAT 270
 QY 108 ThrAspGluArglyserArglyTyrAsnMetSerLeuGlyGluArgAlaValaAla 127
 |||||
 DB 271 ACCGACTCCGCTCGTACCGCAGCTTACCAACGAGCTGCTCCGAGCTGTGCGCGTACAGCC 330
 QY 128 ValArgAspTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleIleSer 146
 |||||
 DB 331 GTTCGTGACGTACTGTGCAACGAGTACGAGTGTGAAGTGTGCGGTAAACGCTGCGGT 390
 QY 147 PheGlyGluArgProIleAlaPheGlyThrAsnGluGlnAlaTrpSerGlnAsnArg 166
 |||||
 DB 391 TACGGGAGTCCCGCGCGCTGCGCGACCAACGCGCAGGCGCGGCTATCAACCGT 450
 QY 167 ArgAlaGlu 169
 |||||
 DB 451 CGCGTTGAA 459

RESULT 9

US-09-267-747-10
 ; Sequence 10, Application US/09267747
 ; Patent No. 6300102
 ; GENERAL INFORMATION:
 ; APPLICANT: Knapp, Bernhard

```

APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Broker, Michael
APPLICANT: Von Specht, Bernd-Ulrich
APPLICANT: Domdey, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-OprI
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/267,747
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/572,447
FILING DATE: 14-DEC-1995
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: linear
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: YES
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
TISSUE TYPE: Serotype 6; ATCC 33354
FEATURE:
NAME/KEY: CDS
LOCATION: 1..483
OTHER INFORMATION: /note="Sequence is coding for oprf
Patent No. 6300102
OTHER INFORMATION: C-terminus"
US-09-267-747-10

Alignment Scores:
Pred. No.: 3.19e-14 Length: 486
Score: 171.50 Matches: 48
Percent Similarity: 46.15% Conservative: 18
Best Local Similarity: 33.57% Mismatches: 67
Query Match: 19.92% Indels: 11
DB: 4 Gaps: 2

US-09-674-779B-2 (1-172) x US-09-267-747-10 (1-486)
QY 37 AlaProThGlyTyrThrGlyValIleTyrThrGlyAlaAlaProLeuValAspAsnsp 56
Db 32 GCTCCGACTCCGACAGCAGCGGCTGCGACAAC-GTGCACAAGTCCCGGACACCCCG 90
QY 57 GluThValIleAlaLeuAlaSerIleuProSerIleu----- 69
Db 91 GCCAAGCTGACGCTTGACGCCAAGCGCTGCCGCGCAAGTGTACGCGTACAG 150

```

```

QY 70 -----ValTyrPheAspPheAspSerAspGluIleIleYsProGlnAlaAlaIleLeu 87
Db 151 CTGACGCTGAAGTTCGACTTCGACAGTCCAAAGTCAAGACAGACGTCGACATC 210
QY 88 AspGluGlnAlaGlnPheLeuThrThrasnGlnThrAlaArgValIleValAlaGlyHis 107
Db 211 AGAAGCTGCCGACTTCATGAGACAGTACCGGTCACCTTCACACCGCTTGAAGGTCAT 270
QY 108 ThrAspGluArgGlySerArgGluTyrAsnMetSerIleuGlyGluArgAlaValAla 127
Db 271 ACCGACTCCGCTCGTACGACGCTTACCAACGAGAGCTGTCCGAGCGCTGCCAAGCC 330
QY 128 ValArgAsnTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleIleSer 146
Db 331 GTTCGTGACGACTGTGTACAGAGTACGCTGTGAGAGTGTCCGCTGACGCTGCGT 390
QY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlnAlaTyrSerGlnAsnArg 166
Db 391 TACGCGAGTCCCGCGCGGCTTGCCGACAAAGCCACCGCTGAGAGCCGCGCTATCAGCT 450
QY 167 ArgAlaGlu 169
Db 451 CGCGTTGAA 459

RESULT 10
US-08-572-447C-14
Sequence 14, Application US/08572447C
Patent No. 595050
GENERAL INFORMATION:
APPLICANT: Knapp, Bernhard
APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Broker, Michael
APPLICANT: Domdey, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-OprI
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/572,447C
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 681 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: linear
HYPOTHETICAL: NO

```


1001 CCAATACCGCAATTACTACAACCCCTTGATTGGTGTATCAATGCGGGTATTCTTAC 1060

```

;
; FEATURE:
; NAME/KEY: CDS

```

```

US-08-467-722A-1
Alignment Scores:
Pred. No.: 1,39e-12 Length: 1720
Score: 165.50 Matches: 47
Percent Similarity: 46.50% Conservative: 26
Best local Similarity: 29.94% Mismatches: 63
Query Match: 19.22% Indels: 21
DB: 3 Gaps: 4

US-09-674-779B-2 (1-172) x US-08-467-722A-1 (1-1720)
QY 35 ProAsnAlaProthrGlyTyr-----ThrGlyValIleTyr 46
    |||||
DB 1001 CCNAATACCGCAATTAACACTACACACCTTGATGTTGATCAATGCGGATATTCTTAC 1060
QY 47 -----ThrGlyValAlaProLeuValAspAsnspGluThrValIysAlaLeuAla 63
    ||| |||||:::||||
DB 1061 CGTTTCGTCACAGCGCAACAGTGTTCACGACACCTCGAATGTAAGTACCAAACTTC 1120
QY 64 SerTysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleTysProGlnAla 83
    ||| ||| ||| ||| ||| |||
DB 1121 AGC---TTAAATTCGATGTAACCTTCCCATTTGGTAAAGCAAACTTAAACCTCAAGA 1177
QY 84 AlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeu 103
    ||| |||||
DB 1178 CAACTTAATTAAGACACGCGTCATGCGCAAAATTTCCAAAGTTAAAGTCGAAAGTACT 1237
QY 104 ValAlaGlyHisThrAspGluArgIysSerArgGluTyrAsnMetSerLeuGlyGluArg 123
    |||||:::|||||
DB 1238 GTTGCTGGTGTACACTACCGATGTTGGTTCGACGCGTTCAACGTAAACTTTCACAGAA 1297
QY 124 ArgAlaValAlaValaArgAsnTyrLeuLeuGlyIysGlyIleAsnGlnAlaSerValGlu 143
    |||||:::|||||
DB 1298 CGTGCAAGATTACGTAGCTAATCTTGTGCTTAAGTGTTGCACGACGCAATTCGA 1357
QY 144 IleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu----- 160
    ::|||:::|||||
DB 1358 GCAACTGTGTACGCTGAACCAAAACCAAGTACGCGCACACTTGTGACCAACTTAAGGT 1411
QY 161 -----AlaTrpSerGlnAsnArgArgAlaGluLeuSer 171
    :::::|||||
DB 1418 CGTAAAGCACTTAATGCTGTGCTTGCTCCAGACCGTGTGTAAGAAATGCA 1468

RESULT 14
US-09-721-017B-786/C
; Sequence 786, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221, 017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546

```

```

FILLING DATE: 30-JAN-1998
PRIOR APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA: PCT/AU98/01023
APPLICATION NUMBER: 10-DEC-1998
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 786:
SEQUENCE CHARACTERISTICS:
LENGTH: 3656 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: FORKPHROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1...3656
US-09-221-017B-786

Alignment Scores:
Pred. No.: 3,49e-09 Length: 3656
Score: 144.00 Matches: 48
Percent Similarity: 37.07% Conservative: 28
Best Local Similarity: 23.41% Mismatches: 83
Query Match: 16.72% Indels: 46
gaps: 5
DB: 4

US-09-674-779B-2 (1-172) x US-09-221-017B-786 (1-3656)

QY      8 AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrglyCysAla---AsnLys 26
       :::          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      2438 AGTGTAGTAGGCTCGTGTTGGCCGTGGCTCTGCTGCCCGGTTCGCAGTAACAAT 2379
QY      27 SerThrSerGlnValMetValAlaProAsnAlaProThrnglyTyrrhglyValIleTyr 46
       ::::          ::::::          -----GCCGAGTAGAGAGGTGCCATTGGT 2331
Db      2378 ATGGCAAAAGCGCGCTTATCGGC-----GCCGAGTAGAGAGGTGCCATTGGT 2331
QY      47 ThrGlyValAlaProLeuValAspAsn-----                    55
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      2330 GCCGAGTAGAGTAGACGTAGCCGGAATACGGCTCGGTGCCATCGTCGACTGCAATC 2271
QY      56 -----AspGluThrValLysAlaLeuAla 63
       ::::          ::::::          -----
Db      2270 GGTTGAGAGCACCGCGGTCTCTCATCGAAGAAGATGACAGCAAGAAAAAGAACCTGAG 2211
QY      64 SerLysLeuProSer-----LeuValTyr 71
       ::::          ::::::          -----
Db      2210 GCCGAGTAGACCGAGTCTAGTACGATTCAGATTCAGACAGTAATAATGACGAGAGGCTATTCTGGTTACT 2151
QY      72 PheASP-----PheaspSeraspGluileLysproGlnAlaAlaAla 85
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      2150 TTTCGATAGCGGATATCTCTTTTCGACAGAACTCCAGACACTGTAGTCCCAACTCAGCACT 2091
QY      86 IleLeuaspGluGlnAlaGlnPheLeuThrTrpAsnGlnThrAlaArgValLeuValAla 105
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      2090 GGCCTGACCAAGTTGGCTCCAACAATGAAACAAAACCAGCAGCAGATTTGTATTCGTA 2033
QY      106 GlyHisThrAspGluArgLysSerArgGluTyrAsnMetSerLeuGlyGluArgArgAla 125
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      2030 GGGCAATACGACATATACGGGTCTCCGACAGATCAACGATCTCTGTGTGAGAGAGCTGCA 1971

```


QY 8 AlaAlaAlaAlaAlaAlaSerValIleuThrPheMetThrLysAlaAlaSer 27
| | | | | | | | | | : : : : :
Db 1003111 GCGCGCAAGGGCGCTTATGACGCGCCTCAACGCGTTGCTGCGGCGGTGAACGTC 10031170

QY 28 ThrSerGlnValMetValAlaPro-----AsnAlaProThr 39
| | | : : : | | |
Db 1003171 ATCGACAGGATTCACGTCGATCCGCTGTGGCATCACTTGATTTCTCAAGTCGGAACCA 1003230

QY 40 GlyYrYrThngLy-----Val 44
: : : : :
Db 1003231 GTTTTCACCGCGAGCGTCGCGCATTCGATTTTGGCCTCAAAAGTCGAAGGACACCGTC 1003290

QY 45 IleYrThrGlyValAlaProLeuValAspAsnAspGluThrValLys---AlaLeuAla 63
| | | | | | | | | | : : : : : | | |
Db 1003291 ACCTTGACCGGAACTGCCCTTCATCCGACCAAGGACGAGTGAAGGCGCGCGCGACC 1003350

QY 64 SerLysIleuProSerLeu----- 69
| | | : : :
Db 1003351 AGCACCTGGCTGACATGTAAATCGTTAAACAATATGAGTTACGGGCGACGACACGCCA 1003410

QY 69 ----- 69
Db 1003411 GGACCCCGGCGCTCCGGGCCCATGTGCGACCTGCAATCAGCATCAATAGCCGTGACGGGT 1003470

QY 70 -----ValYrPheAspPheAspSerAspGluIleLysProGlnAlaAlaAlaIleLeu 87

QY 8 AIAAIAAIAAIAAIAeSerValIeuThrPheMetThrLysAlaAnsSsr 27
Db 1003110 | | | | | : : : : |
GCCGCAGGGCGCCCTTGATACGCGCCTCAACGCGTTGCTTCGCCGGCGTGAACGTC 1003166
QY 28 ThrSerGlnvalMetValAlapro-----AsnAlapProthr 39
Db 1003170 ATCGACCAGAAITCAACGTGATCCGCTGTGCATCATCTGATTTCmTAAGTCGGAGAACA 1003222
QY 40 GLyTYrThrgLy-----Val 44
Db 1003230 GTTTTCACCGCCAGCGtBCCGATTTCTGATTTGGCCTCAAAGTCSAAAAGGACACCGTC 1003289
QY 45 IleTyrrTrgIyValAlaproLeuValAspAsnAspGluthrYalys---AlaleuAla 63
Db 1003290 ACCTTGACCGGAGATCGCCCTTCATCGACACACAAGAGACGAGTGAACGGCCCGCGGACAC 1003349
QY 64 SerlyLeuProSerLeu----- 69

836-500A-1

```

Sequence 1, application US/0836500A
Patent No. 6197929
GENERAL INFORMATION:
APPLICANT: Binz, Hans
APPLICANT: Baussant, Thierry
APPLICANT: Haeuw, Jean-Francois
APPLICANT: Nguyen Ngoc, Thien
TITLE OF INVENTION: Carrier Protein Having an Adjuvant
EFFECT: Immunogenic Complex Containing It, Process for
TITLE OF INVENTION: Their Preparation, Nucleotide Sequence and Vaccines
Patent No. 6197929
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rokey, Milnamow & Katz, Ltd.
STREET: 180 N. Stetson, 2 Prudential Plaza, Suite
STREET: 4700
CITY: Chicago
STATE: Illinois
COUNTRY: U.S.A.
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,500A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Katz, Martin L.
REGISTRATION NUMBER: 25,011
REFERENCE/DOCKET NUMBER: PLEI514P0180US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5400
TELEFAX: 312-616-5460
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1007 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURES:
NAME/KEY: CDS
LOCATION: 1..1007
US-08-836-500A-1
Alignment Scores:
Pred. No.: 5.17e-09 Length: 1007
Score: 136.50 Matches: 35
Percent Similarity: 47.41% Conservative: 20
Best Local Similarity: 30.17% Mismatches: 50
Query Match: 15.85 Indels: 11
DB: Gaps: 2
US-09-674-779B-2 (1-172) x US-08-836-500A-1 (1-1007)
QY 66 LeuProSerLeuValrPhaSpheAspSerAspGluLeuLysProGlnAlaAlaAla 85
DB 616 CTGAAGCTTGACGTTCCTTCAACTTCAACAAGAAGCTCCGTAACCGGAAGTCAAGC 675
QY 86 lIleuAspGluInAlaGlnPheLeuThr-----AsnGlnThrAlaArgValIleu 103
DB 676 GCCTCGATGACGTGCACACGATCAGCTGAGCAAAATGATGCCGAAGACGGTTCCGCTTT 735
QY 104 ValAlaGlyHisThrAspGluArgGlySerArgGluLysIrisMetSerLeuGlyGuar 123
DB 736 GTTCTGGCGTACACCACGACGATCGGTCTCCGAAGCTTACACACGACGAGCTGTGAGAAA 795
QY 124 ArgAlaValAlaValArGAsnTrIleuLeuGlyLysGlyILeaSnGlnAlaSerValGlu 143
DB 796 CGTCTAGTACGCTGTGTGACACTACCTGGTTGCTAAAGCATCCCGCGTGGCAAAATCTCC 855

```

```

Oy      144  lileleserhbeqlgtugluarqprolllealapheslytThrasnlu----- 159
Db      856  GCTTCGGCATGGATGATGCCAACCAGGTACTGCGACACACCTGTGACACAGCGAAGCT 915
Oy      160  -----GuaIatrpserglnAsnaAArgAAlaIuleu 170
Db      916  CGCGCTCCCTGATGATGATGCGCTGCGCTCCGAGTCGCTGAGATGATC 963

RESULT 21
US-08-721-979A-13
: Sequence 13, Application US/08721979A
: Patent No. 6113911
:
GENERAL INFORMATION:
APPLICANT: Binz, Hans
APPLICANT: N'Guyen, Ngoc Thien
APPLICANT: Bausant, Thierry
APPLICANT: Trudel, Michel
TITLE OF INVENTION: PEPTIDE FRAGMENT OF RESPIRATORY
TITLE OF INVENTION: SYNCTIAL VIRUS PROTEIN G, IMMUNOGENIC AGENT, PHARMACEUTICAL
TITLE OF INVENTION: COMPOSITION CONTAINING IT AND PREPARATION PROCESS
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gordon W. Hueschen
STREET: 715 The "H" Bldg., 310 East Michigan
CITY: Kalamazoo
STATE: MI
COUNTRY: USA
ZIP: 49007
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721,979A
FILING DATE: October 4, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 94 04009
FILING DATE: 06-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hueschen, Gordon W.
REGISTRATION NUMBER: 16,157
REFERENCE/DOCKET NUMBER: PF57PCTUS/dln
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-382-0030
TELEFAX: 616-382-2030
INFORMATION FOR SEQ. ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1008 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1008
OTHER INFORMATION: /note= "name : P40"

US-08-721-979A-13

Alignment Scores:
Prod. No.: 5,17e-09 Length: 1008
Score: 136.50 Matches: 35
Percent Similarity: 47.41% Conservative: 20
Best Local Similarity: 30.17% Mismatches: 50
Query Match: 15.85% Indels: 11
DB: 3 Gaps: 2

US-09-674-779B-2 (1-172) x US-08-721-979A-13 (1-1008)

```

```

QY 66 LeuProSerLeuValIyrPheaspPheaspSerAspGluIleIysProGlnAlaAla 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 CTGAAGTCTGACCTGCTGTCACTCAACAAAGCTACCAAGCGAGTCAAGCAG 675
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 IleuAspGluGlnAlaGlnPheLeuThr-----AsnGlnThrAlaArgValLeu 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 GCCTGGATCAGCTGTACACCTCAGCAGCAACATGATCCCAAGACGCTCCGCTGT 735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 104 ValAlaGlyHisThrAspGluArgGlySerArgGlyIleAsnMetSerLeuGlyGly 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 736 GTTCTGGCTACACCGACCGCATCGCTCCGAAGCTTACACACGACGCTGTCTGAGAA 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 ArgAlaValAlaValArgAsnTyrLeuLeuGlyIleAsnGlnAlaSerValGlu 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 796 CGTCTCAGCTCCGTTGTTGACTACTGCTTAAAGGATCCGCGTGGCAAAATCTCC 855
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 IleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGlu----- 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 GCTCGGCGATGGGTAATCCAAACCGGTTACTGCGCAACACCTGTGACAAAGCTGAAAGCT 915
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 -----GluAlaIlePserGlnAsnArgAlaGluLeu 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 916 CCGCGTGCCTGATGATGCTGCTGCGATCGCTGCTGATGAGATC 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 22
; Sequence 13, Application US/09654289
; Patent No. 6410030
; GENERAL INFORMATION:
; APPLICANT: Binz, Hans
; APPLICANT: N'Guyen, Ngoc Thien
; APPLICANT: Baussant, Thierry
; APPLICANT: Trudel, Michel
; TITLE OF INVENTION: PEPTIDE FRAGMENT OF RESPIRATORY
; TITLE OF INVENTION: SYNCTIAL VIRUS PROTEIN G, IMMUNOGENIC AGENT, PHARMACEUTICAL
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gordon W. Hueschen
; STREET: 715 Tne "H" Bldg., 310 East Michigan
; CITY: Kalamazoo
; STATE: MI
; COUNTRY: USA
; ZIP: 49007
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/654,289
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/721,979
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hueschen, Gordon W.
; REGISTRATION NUMBER: 16,157
; REFERENCE/DOCKET NUMBER: P57PCTUS/41n
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-382-0030
; TELEFAX: 616-382-2030
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1008 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:

```

```

; NAME/REV: CDS
; LOCATION: 1..1008
; OTHER INFORMATION: /note= "name : P40"
; US-09-654-289-13

Alignment Scores:
Pred. No.: 5.17e-09 Length: 1008
Score: 136.50 Matches: 35
Percent Similarity: 47.41% Conservative: 20
Best Local Similarity: 30.17% Mismatches: 50
Query Match: 15.85% Indels: 11
DB: 4 Gaps: 2

US-09-674-779b-2 (1-172) x US-09-654-289-13 (1-1008)

QY 66 LeuProSerLeuValIyrPheaspPheaspSerAspGluIleIysProGlnAlaAla 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 CTGAAGTCTGACCTGCTGTCACTCAACAAAGCTACCAAGCGAGTCAAGCAG 675
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 IleuAspGluGlnAlaGlnPheLeuThr-----AsnGlnThrAlaArgValLeu 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 GCCTGGATCAGCTGTACACCTCAGCAGCAACATGATCCCAAGACGCTCCGCTGT 735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 104 ValAlaGlyHisThrAspGluArgGlySerArgGlyIleAsnMetSerLeuGlyGly 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 736 GTTCTGGCTACACCGACCGCATCGCTCCGAAGCTTACACACGACGCTGTCTGAGAA 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 ArgAlaValAlaValArgAsnTyrLeuLeuGlyIleAsnGlnAlaSerValGlu 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 796 CGTCTCAGCTCCGTTGTTGACTACTGCTTAAAGGATCCGCGTGGCAAAATCTCC 855
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 IleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGlu----- 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 GCTCGGCGATGGGTAATCCAAACCGGTTACTGCGCAACACCTGTGACAAAGCTGAAAGCT 915
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 -----GluAlaIlePserGlnAsnArgAlaGluLeu 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 916 CCGCGTGCCTGATGATGCTGCTGCGATCGCTGCTGATGAGATC 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 23
; Sequence 14, Application US/08129719
; Patent No. 5556755
; GENERAL INFORMATION:
; APPLICANT: Timothy F. Murphy
; TITLE OF INVENTION: Vaccine For Branhamella catarrhalis
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
; STREET: 1800 One Met Plaza
; CITY: Buffalo
; STATE: New York
; COUNTRY: United States
; ZIP: 14203-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
; SOFTWARE: Wordperfect for Windows 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/129,719
; FILING DATE: September 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Nelson, M. Bud
; REGISTRATION NUMBER: 35,300
; REFERENCE/DOCKET NUMBER: 11520.0050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 856-4000
; TELEFAX: (716) 849-0349
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1727 nucleotides
; TYPE: nucleic acid

```

```
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOHETICAL: yes
ORIGINAL SOURCE:
ORGANISM: Branchamella catarrhalis
STRAIN: 25240
IMMEDIATE SOURCE:
LIBRARY: genomic
CLONE: EMBL3 clone 5
CHROMOSOME/SEGMENT: pcd1
FEATURE:
LOCATION: CD gene region
IDENTIFICATION METHOD: by experiment
NAME/KEY: signal sequence of encoded protein
LOCATION: -26 to -1
US-08-129-719-14

Alignment Scores:
Pred. No.: 1.36e-08 Length: 1727
Score: 136.00 Matches: 39
Percent Similarity: 49.07% Conservative: 14
Best Local Similarity: 36.11% Mismatches: 43
Query Match: 15.80% Indels: 12
DB: 1 Gaps: 5

US-09-674-779b-2 (1-172) x US-08-129-719-14 (1-1727)
QY 70 ValTyrPheaspheaspseraspgluileysprogin-----AlaIaIaIa 85
DB 1075 GTATCTCTTGATGATGATTAATCAATCAATCAACCAACCAATGAGTGTGTAAG 1134
QY 86 IleleuaspqluglnaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 105
DB 1135 GTTCCTGCGCAATGCGTGAATTC-----CCAAATGCAACTGCA-----ACCATGAA 1182
QY 106 GlyHisThrAspGlu-----ArgGlySerArgGluTyrAsnMetSerLeuGlyGlu 122
DB 1183 GGTCAACGATCAGCGATTCAGCAACGCTCAAGTGCAGCAACGCTCTATCTGAA 1242
QY 123 ArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSer 141
DB 1243 GCTCGTCTGTAATGCTTAATCAATGCTATCAAGCAATTTGGTATCGCTCCAAACCGC 1302
QY 142 ValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAla 161
DB 1303 CTAAATGCAAGTTGGTTATGCGCTTGATGCTCTATCGCTCAAACTACTACTGTAAGGT 1362
QY 162 TrpSerGlnAsnArgArgAlaGlu 169
DB 1363 AAAGCGATGAACCGCTGTAAGAA 1386

RESULT 24
US-08-306-871-14
; Sequence 14, Application US/08306871
; Patent No. 5712118
; GENERAL INFORMATION:
; APPLICANT: Timothy F. Murphy
; TITLE OF INVENTION: Vaccine For Branchamella catarrhalis
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
; STREET: 1800 One Mt Plaza
; CITY: Buffalo
; STATE: New York
; COUNTRY: United States
; ZIP: 14203-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
; SOFTWARE: Wordperfect for Windows 5.1
; CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/08/306,871
FILING DATE: 20-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/129,719
FILING DATE: September 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Nelson, M. Bud
REGISTRATION NUMBER: 35,300
REFERENCE/DOCKET NUMBER: 11520.0053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 856-4000
TELEFAX: (716) 849-0349
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1727 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOHETICAL: yes
ORIGINAL SOURCE:
ORGANISM: Branchamella catarrhalis
STRAIN: 25240
IMMEDIATE SOURCE:
LIBRARY: genomic
CLONE: EMBL3 clone 5, SUBCLONE: pcd1
FEATURE:
LOCATION: CD gene region
IDENTIFICATION METHOD: by experiment
NAME/KEY: signal sequence of encoded protein
LOCATION: -26 to -1
US-08-306-871-14

Alignment Scores:
Pred. No.: 1.36e-08 Length: 1727
Score: 136.00 Matches: 39
Percent Similarity: 49.07% Conservative: 14
Best Local Similarity: 36.11% Mismatches: 43
Query Match: 15.80% Indels: 12
DB: 1 Gaps: 5

US-09-674-779b-2 (1-172) x US-08-306-871-14 (1-1727)
QY 70 ValTyrPheaspheaspseraspgluileysprogin-----AlaIaIaIa 85
DB 1075 GTATCTCTTGATGATGATTAATCAATCAATCAACCAACCAATGAGTGTGTAAG 1134
QY 86 IleleuaspqluglnaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 105
DB 1135 GTTCCTGCGCAATGCGTGAATTC-----CCAAATGCAACTGCA-----ACCATGAA 1182
QY 106 GlyHisThrAspGlu-----ArgGlySerArgGluTyrAsnMetSerLeuGlyGlu 122
DB 1183 GGTCAACGATCAGCGATTCAGCAACGCTCAAGTGCAGCAACGCTCTATCTGAA 1242
QY 123 ArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSer 141
DB 1243 GCTCGTCTGTAATGCTTAATCAATGCTATCAAGCAATTTGGTATCGCTCCAAACCGC 1302
QY 142 ValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAla 161
DB 1303 CTAAATGCAAGTTGGTTATGCGCTTGATGCTCTATCGCTCAAACTACTACTGTAAGGT 1362
QY 162 TrpSerGlnAsnArgArgAlaGlu 169
DB 1363 AAAGCGATGAACCGCTGTAAGAA 1386

RESULT 25
US-08-569-959-14
; Sequence 14, Application US/08569959
; Patent No. 5725862
; GENERAL INFORMATION:
; APPLICANT: Timothy F. Murphy
```

;; TITLE OF INVENTION: Vaccine For Branhamella catarrhalis
;; NUMBER OF SEQUENCES: 52
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
;; STREET: 1800 One Mt Plaza
;; CITY: Buffalo
;; STATE: New York
;; COUNTRY: United States
;; ZIP: 14203-2391
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
;; SOFTWARE: Wordperfect for Windows 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/569,959
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: U.S. 08/129,719
;; FILING DATE: September 29, 1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Nelson, M. Bud
;; REGISTRATION NUMBER: 35,300
;; REFERENCE/DOCKET NUMBER: 11520.0053
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (716) 856-4000
;; TELEFAX: (716) 849-0349
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1727 nucleotides
;; TYPE: nucleic acid
;; STRANDEDNESS: single-stranded
;; TOPOLOGY: linear
;; MOLECULE TYPE: genomic DNA
;; HYPOTHEICAL: yes
;; ORIGINAL SOURCE:
;; ORGANISM: Branhamella catarrhalis
;; STRAIN: 25240
;; IMMEDIATE SOURCE:
;; LIBRARY: genomic
;; CLONE: EMBL3 clone 5
;; FEATURE:
;; LOCATION: CD gene region
;; IDENTIFICATION METHOD: by experiment
;; NAME/KEY: signal sequence of encoded protein
;; LOCATION: -26 to -1
US-08-569-959-14
Alignment Scores:
Pred. No.: 1.36e-08 Length: 1727
Score: 136.00 Matches: 39
Percent Similarity: 49.07% Conservative: 14
Best Local Similarity: 36.11% Mismatches: 43
Query Match: 15.80% Indels: 12
DB: 1 Gaps: 5
US-09-674-779B-2 (1-172) x US-08-569-959-14 (1-1727)
QY 70 ValTyrPheAspPheAspSerAspGluIleLysProGln-----AlaAlaAla 85
Db 1075 GTATTCTTGTGATTGATTAATCAATCAATCAACCAATACCGTGAAGAAGTGTCTAAG 1134
QY 86 ILeLeuAspGluGlnAlaGlnPheLeuThrAsnGlnTrpAlaArgValLeuValAla 105
Db 1135 GTTGTGGCGCAATCGCGATTC-----CCAATGCAACTGCA-----ACCATTGAA 1182
QY 106 GlyHISTrpAspGln-----ArgGlySerArgIuTrpAsnMetSerLeuGlyGlu 122
Db 1183 GGTCCAGCATCAGCGATTCAAGCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGTGA 1242
QY 123 ArgArgAlaValAlaValAlaArgAsnTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSer 141
||||| |||||||:|:| |||

Db 1243 GCTCGCTAATGCTGTTAATCAATCTATCAACGAATTTGTATCCCTCCAAACCGC 1302
QY 142 ValGluIleIleSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGluAla 161
Db 1303 CTAATGCACTGTGTTATGCTTGTGATCGTTCATCGCTCAATATCTACTGCTGAAGT 1362
QY 162 TrpSerGlnAsnArgArgAlaGlu 169
Db 1363 AAAGCATGAACCGCTGTAGAA 1386
RESULT 26
US-08-469-260A-80
; Sequence 80, Application US/08469260A
; Patent No. 6451578
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAWSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MOERHOFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUIK
; APPLICANT: ISA K. MUSHAMMAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,260A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4268 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-260A-80
Alignment Scores:
Pred. No.: 0.192 Length: 4268
Score: 87.50 Matches: 38
Percent Similarity: 41.04% Conservative: 17
Best Local Similarity: 26.36% Mismatches: 48
Query Match: 10.16% Indels: 31
DB: 4 Gaps: 6
US-09-674-779B-2 (1-172) x US-08-469-260A-80 (1-4268)

```

OY 3 LeuHISileglnlleAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGly 22
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 1686 CTGGCTGTGGCGGTGGAGCTATGAGCTTATGACCATGACACTTTGGCGCCACT 1745
OY 23 CysAlaAsnLysSerThrSerglnValMetValAlaProAsnAlaProThrGlyTyrThr 42
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 1746 TGTGTGGCGGCTGTGCTGCTATTATCATCATGTC-----CTTACCGGTGCTACT 1793
OY 43 GlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeu 62
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 1794 -----GTGCCCCAGGTGTGACGAGAAATCGTGAGAGAGT 1835
OY 63 AlaSerLysLeuProSerLeuValIlyrPheAspPheAspSerAspGluIleLysProGln 82
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 1836 GCATCATTCATTCC-----TTGAGGCCCATG 1862
OY 83 AlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgVal 102
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 1863 GTTGCTGCATTT---GACAACTGGAAGAGTACATACACACACA----- 1901
OY 103 LeuValAlaGlyHisThrAspGluArgGlySerArgGlyTyrAsnMetSerLeuGlyGlu 122
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 1902 ---ACTAGTCCTTGCATTTGGAACCCGCTTGAATAAATTAAACACCTTCTTGTGGCCT 1958
OY 123 ArgArgAla-----ValAlaValAlaArgAsnTyrLeuLeuGly 134
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 1959 CATGCAGCTACAACTCTGCTCATCATGAGATTTGCTGTGGC 2000

```

RESULT 27

```

US-08-469-260A-11
; Sequence 11, Application US/08469260A
; Patent No. 6451578

```

```

; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAM J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAMSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MUEHRHOF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUIK
; APPLICANT: ISA K. MUSHAWAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,260A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMBESKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 11:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 8912 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-260A-11

```

```

Alignment Scores:
Pred. No.: 0.583
Score: 87.50
Percent Similarity: 41.04%
Best Local Similarity: 28.36%
Query Match: 10.16%
DB: 4 Gaps: 6
Matches: 38
Conservative: 17
Mismatch: 48
Indels: 31

```

```

US-09-674-779b-2 (1-172) x US-08-469-260A-11 (1-8912)

```

```

OY 3 LeuHISileglnlleAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGly 22
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 4953 CTGGCTGTGGCGGTGGAGCTATGAGCTTATGACCATGACACTTTGGCGCCACT 5012
OY 23 CysAlaAsnLysSerThrSerglnValMetValAlaProAsnAlaProThrGlyTyrThr 42
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 5013 TGTGTGGCGGCTGTGCTGCTATTATCATCATGTC-----CTTACCGGTGCTACT 5060
OY 43 GlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeu 62
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 5061 -----GTGCCCCAGGTGTGACGAGAAATCGTGAGAGAGTGT 5102
OY 63 AlaSerLysLeuProSerLeuValIlyrPheAspPheAspSerAspGluIleLysProGln 82
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 5103 GCATCATTCATTCC-----TTGAGGCCCATG 5129
OY 83 AlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgVal 102
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 5130 GTTGCTGCATTT---GACAACTGGAAGAGTACATACACACA----- 5168
OY 103 LeuValAlaGlyHisThrAspGluArgGlySerArgGlyTyrAsnMetSerLeuGlyGlu 122
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 5169 ---ACTAGTCCTTGCATTTGGAACCCGCTTGAATAAATTAAACACCTTCTTGTGGCCT 5225
OY 123 ArgArgAla-----ValAlaValAlaArgAsnTyrLeuLeuGly 134
   ||| : : : : : ||| ||| ||| : : : : : ||| |||
Db 5226 CATGCAGCTACAACTCTGCTCATCATGAGATTTGCTGTGGC 5267

```

RESULT 28

```

US-08-639-857-32
; Sequence 32, Application US/08639857
; Patent No. 5955318

```

```

; GENERAL INFORMATION:
; APPLICANT: Simons, J. N.
; APPLICANT: Desai, S. M.
; APPLICANT: Mushawar, I. K.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR CONTROLLING THE
; TITLE OF INVENTION: TRANSLATION OF HEPATITIS GB PROTEINS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,857
; FILING DATE:
; CLASSIFICATION: 435

```



```

? ADDRESS: ABBOTT LABORATORIS D377/AP6D
? STREET: 100 ABBOTT PARK ROAD
? CITY: ABBOTT PARK
? STATE: IL
? COUNTRY: USA
? ZIP: 60064-3500
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/469,260A
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/424,550
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: FOREMSKI, PRISCILLA E.
? REGISTRATION NUMBER: 33,207
? REFERENCE/DOCKET NUMBER: 5527.PC.01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 708-937-6365
? TELEFAX: 708-938-2623
? INFORMATION FOR SEQ ID NO: 390:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 9143 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
US-08-469-260A-390

Alignment Scores:
Pred. No.:          0.606      Length:       9143
Score:              87.50     Matches:        38
Percent Similarity: 41.04%    Conservative: 17
Best Local Similarity: 28.36% Mismatches:   46
Query Match:         10.16%   Indels:        31
                        Gaps:           6

US-09-674-779B-2 (1-172) x US-08-469-260A-390 (1-9143)
QY      3 LeuHsiIleGInIIlaAlaAlaAlaAlaLeuSerValLeuThrPhemeTrhcy 22
DB      5141 CTCGGTGTGGCGCTTGAGATGGCTATGCTTATCTTAGCCATTGACACTTTGGCGGCAC 5200
QY      23 CysAlaAsnLysSerThrSercInValMetValAlaProAsnAlaProThrGIYTYrThr 42
DB      5201 TGTGTGCCGCGTTCGCTGCTATTATTAACATCACTC-----CCATCCGGGTGCTACT 5248
QY      43 GlyValIleTyrrThnglyValAlaProLeuValAspAsnaSpGIunThrValLysAlaLeu 62
DB      5249 -----GTCGCCCACTGGTTGACGAAGAAGAAATCGTGAGAGAGTGT 5290
QY      63 AlaSerLysIeuProSerLieuValTyrrPheaspPheaspSerAspGIunIleLysProGln 82
DB      5291 GCATCATTCATTCATCC-----TTGAGAGCCCATG 5317
QY      83 AlalAlaIleLeuaspGIunGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgVal 102
DB      5318 GTTCTCTCAATC---GATAAGCTGAGAGATACAATACATCACCA-----5356
QY      103 LeuValAlaGlyHisThrAspGIunArgLySerArgGIunTyrrAsnMetSerLeuGIun 122
DB      5357 ---ACTAGTCCTTTCACATTTGGAACCGCCCTTGAAAACACTTAACACACTTTCYTGGGCT 5413
QY      123 ArgArgAla-----ValAlaValArgAsnTyrrLeuLeugly 134
DB      5414 CATGCACTACAAATCCTTGTCTATCATGAGATATTGCTGTGGC 5455
```

```

Db      5141 CTGGCTGTTGGCGTTGGAGTGGCGCTATGGCTTATCTAGTACGATTGACACTTTTGGCGGCACT 5200
Qy      23  CysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThr 42
      |||  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      5201 TGTGTGGCGGCTTGCTGCTGTCTATTATCATCGATC-----CCATCCGGTGTCTACT 5248
Qy      43  GlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValIleValAlaLeu 62
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      5249 -----GTCCGCCCGAGTGGTGGACCAAAABAATCGGAGGAGGTG 5290
Qy      63  AlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGln 82
      |||||  ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      5291 GCATCATTCATTCCTCC-----TTGGAGGCCATG 5317
Qy      83  AlaAlaIleIleLeuAspGlnGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgVal 102
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      5318 GTTCTCTCAATC--GATTAACGTCAAGATACATACATACACA----- 5356
Qy      103  LeuValAlaGlnHisThrAspGluThrArgGlySerArgGluTyrAsnMetSerLeuGln 122
      ::  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      5357 ---ACTAGTCTCTTACATTCGAAACCGCCCTTGAAAAAATTATACACCTTCTTG6GCGT 5413
Qy      123  ArgArgAla-----ValAlaValArgAsnTyrLeuLeuGly 134
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      5414 CATGCACGCTACATCCTTGCTATCATAGATATGCTGTGCG 5455

RESULT 31
US-09-556-877-171
; Sequence 171, Application US/09556877
; Patent No. 6432916
; GENERAL INFORMATION:
; APPLICANT: Probst, Peter
; APPLICANT: Bhalla, Ajay
; APPLICANT: Skelky, Yasir
; APPLICANT: Fling, Steve
; APPLICANT: Maisonneuve, Jeff
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C5
; CURRENT APPLICATION NUMBER: US/09/556,877
; CURRENT FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 171
; LENGTH: 2895
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-556-877-171

Alignment Scores:
Pred. No.: 0.81 Length: 2895
Score: 81.00 Matches: 39
Percent Similarity: 37.58% Conservative: 23
Best Local Similarity: 23.64% Mismatches: 75
Query Match: 9.41% Indels: 28
DB: 4 Gaps: 7

US-09-674-779B-2 (1-172) x US-09-556-877-171 (1-2895)
Qy      8  AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLysSer 27
      |||  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      829  GCTTCTGATGAGGAGCACTTAAAGTAACTACTCCCTAGATGTTACAGGCAATCGT--- 885
Qy      28  ThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThr 47
      ::|||  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      886  ---GTAAGATATCTTTTATGTGACATATCACAAAAAATTAATGAGGAGCTTTTACGCT 942
Qy      48  GlyValAlaProLeuValAspAsnAspGluThr-----ValIleAlaLeuAlaSerLys 65
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      943  CCTGTAGTAACTACTGATATATGCGCTTACTTATTAACAATATCGCCATATAT 100
Qy      66  LeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaAla 85
      ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

Db 1003 AAGGGGGCGCTATCTATATAGACGAGAACAGTAAC-----TCCAAAT 1047
QY 86 ILeuAspGluGlnAlaGlnPheLeuThrThraSngInThraAlaArgValLeuVala 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1048 TCTGCCGACCGGCATGCTATTTATTTAATGAATAATTTGACTAATTAAGTAAGTCA 1107
QY 106 GLYHISThrAspGluArgGlySerArgGluTyraSmetSerLeuGlyGluArgArgLa 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 AATGTACACAGTACGACTAATCTCTCT-----AGAGAAATGCA 1149
QY 126 ValAlaValArgAsn-----TyrLeuLeuGlyLysGlyIleasngIn----- 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1150 ATACAGTAGACAGCTCTGCTGTAATTTCTATTATAGACGAGGAGGAGCAAAATTTA 1209
QY 140 -----AlaSerValGluIleIleSerPheGlyGlyGluArgProIleAlaPheGly 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1210 ATTTTATGATCCCTATGTAAGTTAGCAATGCAAGG-----GTCTGTGTCTC 1257
QY 157 ThrAsngGluAla 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1258 TTCATTAAGGAAGCT 1272

RESULT 32
US-09-620-412C-171
; Sequence 171, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620.412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 171
; LENGTH: 2895
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-620-412C-171

Alignment Scores:
Pred. No.: 0.81 Length: 2895
Score: 81.00 Matches: 39
Percent Similarity: 37.58% Conservative: 23
Best Local Similarity: 23.64% Mismatches: 75
Query Match: 9.41% Indels: 28
Gaps: 7

US-09-674-779B-2 (1-172) x US-09-620-412C-171 (1-2895)
QY 8 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 27
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 GCTTCTGATGAGAGCAATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 885
QY 28 ThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThr 47
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 ---GGTAGATCTTTTATTAGTACAAATATACAAAAAATTTATGCGAGCTATTTCGCT 942
QY 48 GlyAlaAlaProLeuValAlaAspAsnAspGluThr-----ValLysAlaLeuAlaSerLys 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 CCTGTACTTACCTAGTGAATGAGTAATGCGCTACTTATTAACAATATCGCAATAT 1002
QY 66 LeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaAla 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 AAGGGGGCGCTATCTATATAGACGAGAACAGTAAC-----TCCAAAT 1047
QY 86 ILeuAspGluGlnAlaGlnPheLeuThrThraSngInThraAlaArgValLeuVala 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1048 TCTGCCGACCGGCATGCTATTTATTTAATGAATAATTTGACTAATTAAGTAAGTCA 1107
QY 106 GLYHISThrAspGluArgGlySerArgGluTyraSmetSerLeuGlyGluArgArgLa 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1108 AATGTACACAGTACGACTAATCTCTCT-----AGAGAAATGCA 1149
QY 126 ValAlaValArgAsn-----TyrLeuLeuGlyLysGlyIleasngIn----- 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1150 ATACAGTAGACAGCTCTGCTGTAATTTCTATTATAGACGAGGAGGAGCAAAATTTA 1209
QY 140 -----AlaSerValGluIleIleSerPheGlyGlyGluArgProIleAlaPheGly 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1210 ATTTTATGATCCCTATGTAAGTTAGCAATGCAAGG-----GTCTGTGTCTC 1257
QY 157 ThrAsngGluAla 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1258 TTCATTAAGGAAGCT 1272

RESULT 33
US-09-556-877-183
; Sequence 183, Application US/09556877
; Patent No. 6432916
; GENERAL INFORMATION:
; APPLICANT: Probst, Peter
; APPLICANT: Bhalla, Ajay
; APPLICANT: Skeiky, Yasir
; APPLICANT: Fling, Steve
; APPLICANT: Matsonneuve, Jeff
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C5
; CURRENT APPLICATION NUMBER: US/09/556.877
; CURRENT FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 183
; LENGTH: 2934
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-556-877-183

Alignment Scores:
Pred. No.: 0.826 Length: 2934
Score: 81.00 Matches: 39
Percent Similarity: 37.58% Conservative: 23
Best Local Similarity: 23.64% Mismatches: 75
Query Match: 9.41% Indels: 28
Gaps: 7

US-09-674-779B-2 (1-172) x US-09-556-877-183 (1-2934)
QY 8 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 27
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 868 GCTTCTGATGAGAGCAATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 924
QY 28 ThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThr 47
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 ---GGTAGATCTTTTATTAGTACAAATATACAAAAAATTTATGCGAGCTATTTCGCT 981
QY 48 GlyAlaAlaProLeuValAlaAspAsnAspGluThr-----ValLysAlaLeuAlaSerLys 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 982 CCTGTAGTTACCTAGTGAATGAGTAATGCGCTACTTATTAACAATATCGCAATAT 1041
QY 66 LeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaAla 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1042 AAGGGGGCGCTATCTATATAGACGAGAACAGTAAC-----TCCAAAT 1086
QY 86 ILeuAspGluGlnAlaGlnPheLeuThrThraSngInThraAlaArgValLeuVala 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1087 TCTGCCGACCGGCATGCTATTTATTTAATGAATAATTTGACTAATTAAGTAAGTCA 1146
QY 106 GLYHISThrAspGluArgGlySerArgGluTyraSmetSerLeuGlyGluArgArgLa 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1147 AATGTACACAGTACGCTAATCTCTCT-----AGAGAAATGCA 1188
QY 126 ValAlaValArgAsn-----TyrLeuLeuGlyLysGlyIleasngIn----- 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-674-779B-2 (1-172) x US-09-674-779B-1 (1-80161)
Alignment Scores:
Pred. No.: 359 Length: 80161
Score: 77.50 Matches: 32
Percent Similarity: 33.97% Conservative: 21
Best Local Similarity: 20.51% Mismatches: 48
Query Match: 9.00% Indels: 55
Gaps: 5
DB: 3
US-09-674-779B-2 (1-172) x US-09-674-779B-1 (1-80161)
QY 29 SerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGly 48
Db 5864 ACCGGAGGTGATGTCGACCTCTTGGG-----GCAGGTGTGTCACGCGCGG 5911
QY 49 ValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSer 68
Db 5912 CTCGACACGACGATCCGCTCCACGTCACGACGCA----- 5947
QY 69 LeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAsp 88
Db 5948 -----CTCGAT 5953
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
Db 5954 GATGGAGCCAAATTCAGGCTGCGGTCAACACGACGACCCCTGCGCGGTGATGTG 6013
QY 109 AspGluArg----- 111
Db 6014 ATGGCAGGATAGAGCTGACCGTTGTCAGGTGCGCGAAGGTTCGCGTCAGCCGTCGCCG 6073

QY 112 -----GlySerAlaGluTyrAsnMetSerLeuGlyLysArgAlaVal----- 126
Db 6074 GTAGGTGGATCCACCCGATCACAGTGTCTTCGACGAGGACAGGCTGCTCTTTCG 6133
QY 127 AlaValArgAsnTyrLeuLeuGlyLysGlyTyrLeuGlnAlaSerValGluIleLeuSer 146
Db 6134 GATCTCCGCAATTCGTCAGGCGGAGAGGGGTGCGCAGGT-----GTGCGCCAG 6184
QY 147 PheGlyGluGlnArgProIleAlaPheGlyTyrAsnGluLysAlaTrp 162
Db 6185 CATGATGCTGCGCTCCGCTCCGTCGACGCGGCGCTTATCGGCTCTG 6232
RESULT 37
US-09-370-700-1
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
EARLIER FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1
Alignment Scores:
Pred. No.: 359 Length: 80161
Score: 77.50 Matches: 32
Percent Similarity: 33.97% Conservative: 21
Best Local Similarity: 20.51% Mismatches: 48
Query Match: 9.00% Indels: 55
Gaps: 5
DB: 4
US-09-674-779B-2 (1-172) x US-09-370-700-1 (1-80161)
QY 29 SerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGly 48
Db 5864 ACCGAGGTGATGTCGACCTCTTGGG-----GCAGGTGTGTCACGCGCGG 5911
QY 49 ValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSer 68
Db 5912 CTCGACACGACGATCCGCTCCACGTCACGACGCA----- 5947
QY 69 LeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAsp 88
Db 5948 -----CTCGAT 5953
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
Db 5954 GATGGAGCCAAATTCAGGCTGCGGTCAACACGACGACCCCTGCGCGGTGATGTG 6013
QY 109 AspGluArg----- 111
Db 6014 ATGGCAGGATAGAGCTGACCGTTGTCAGGTGCGCGAAGGTTCGCGTCAGCCGTCGCCG 6073
QY 112 -----GlySerAlaGluTyrAsnMetSerLeuGlyLysArgAlaVal----- 126
Db 6074 GTAGGTGGATCCACCCGATCACAGTGTCTTCGACGAGGACAGGCTGCTCTTTCG 6133


```

QY 58 -----ThrValIysAlaLeuAlaSerIysLeuProSer 68
Db 33316 AGGAGTCTGCGCCGACATGACATGAGTAATGAAATGCGCATTAACCGCGTCA 35375
QY 69 LeuValIYrPheAspPheAspSerAspGluIleIysProGlnAlaAlaIleLeuAsp 88
Db 35376 ATGCTCTTCCCTGAAGCCTCAGATATATGAAAGCAGACGCGCGCGC-----CTGACA 35429
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
Db 35430 GAGCAGAACCGCGTGATT----- 35447
QY 109 AspGluArgGlySerArgGluIYrAsnMetSerLeuGlyGluArgArgAlaValAlaVal 128
Db 35448 GATGAACAGCGCCAGTCGGGTGAATCCCTG-----CAGGAAAAAGCGCATCCATT 35498
QY 129 ArgAsnTYrLeuLeuGlyIYsGlyIleAsnGlnAlaSerValGluIleIleSerPheGly 148
Db 35499 CAGAGTGTGCTTCCCGCGCTG----- 35519
QY 149 GluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166
Db 35520 GAAGACCGTCGTGGCGCTTA--ATTGCTACAGCAGCGGAGCAGATTAAG 35570

RESULT 40
US-09-453-702B-137/C
; Sequence 137, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-Dec-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296, 95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48908
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 137:
US-09-453-702B-137
Alignment Scores:

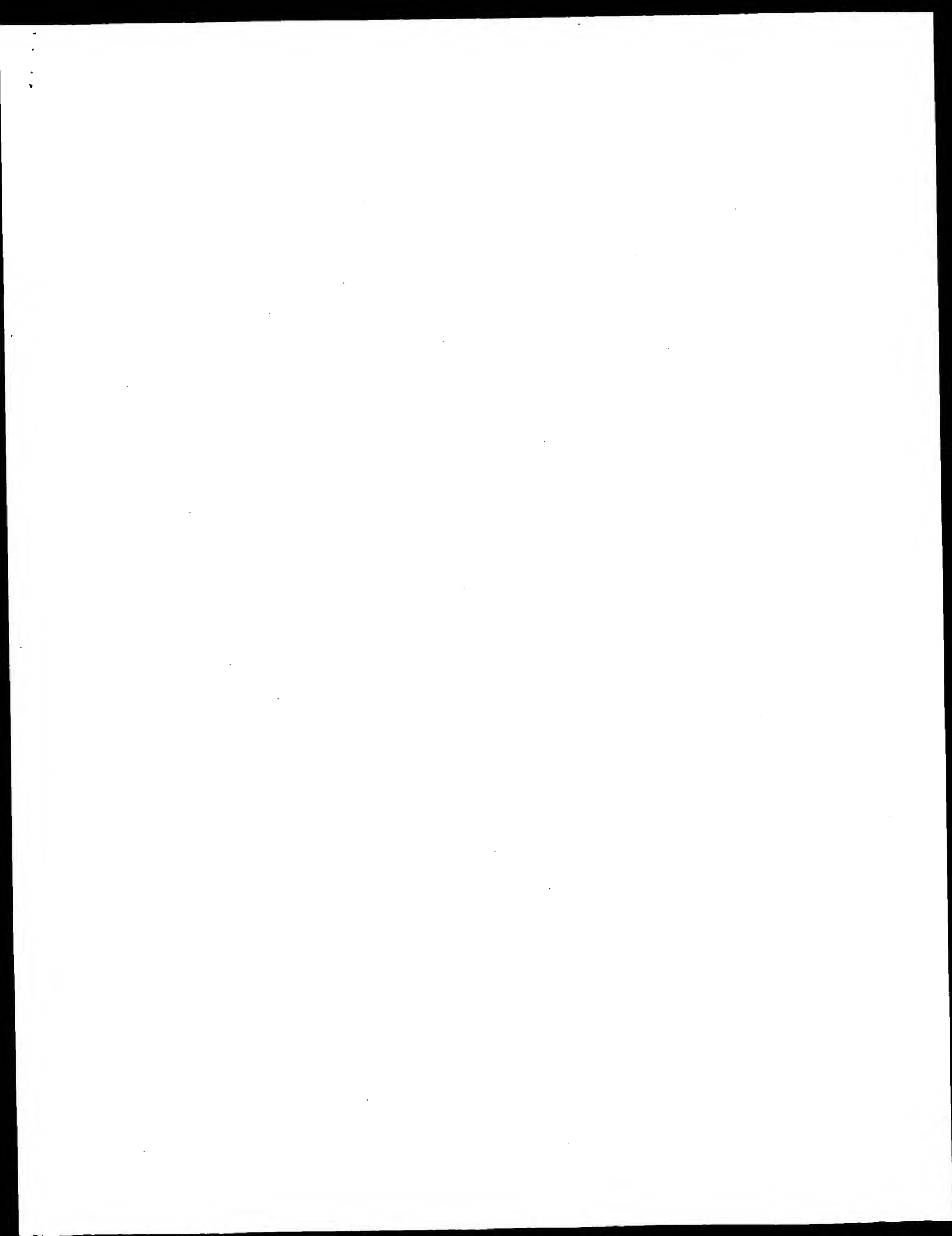
```

```

Pred. No.: 507
Score: 74.00
Percent Similarity: 39.89%
Best Local Similarity: 22.47%
Query Match: 8.59%
DB: 4
Gaps: 8
US-09-674-779b-2 (1-172) x US-09-453-702B-137 (1-48908)
QY 3 LeuHisIleGlnIleAlaAlaAlaAlaAlaLeuSerValIleuThrPheMetThrGly 22
Db 11825 CTGAACAGAAATATTTGCTGCCAGAGCGCCCGACATGCGTGAACAGTACACGCGC 11766
QY 23 CysAlaAsnIYsSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTYrThr 42
Db 11765 GTGGGCTCAGCTGATGATGAGCGGTGGCGCTG-----GGCTGTGTGTGGCTGACCC 11715
QY 43 GlyValIleIleTYrThrGlyValAla-----ProLeuValAspAsnAspGlu----- 57
Db 11714 GGACTGTGATGCTGTGGGGCTGCAGCATGTGTACACGCTGTACAGATCAGACGAGCGCC 11655
QY 58 -----ThrValIysAlaLeuAlaSerIysLeuProSer 68
Db 11654 AGGAGTCTGCGCCGACATGACATGACATGAAATGCGCATTAACCGCGTCA 11595
QY 69 LeuValIYrPheAspPheAspSerAspGluIleIysProGlnAlaAlaIleLeuAsp 88
Db 11594 ATGCTCTTCCCTGAAGCCTCAGATATATGAAAGCAGAACCGCGCGC-----CTGACA 11541
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
Db 11540 GAGCAGAACCGCGTGATT----- 11523
QY 109 AspGluArgGlySerArgGluIYrAsnMetSerLeuGlyGluArgArgAlaValAlaVal 128
Db 11522 GATGAACAGCGCCAGTCGGGTGAATCCCTG-----CAGGAAAAAGCGCATCCATT 11472
QY 129 ArgAsnTYrLeuLeuGlyIYsGlyIleAsnGlnAlaSerValGluIleIleSerPheGly 148
Db 11471 CAGGATGTGCTTCCCGCGCTG----- 11451
QY 149 GluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166
Db 11450 GAAGACCGTCGTGGCGCTTA--ATTGCTACAGCAGCGGAGCAGATTAAG 11400

Search completed: July 6, 2003, 13:40:31
Job time : 1768 secs

```



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:52:45 ; Search time 139 Seconds

(without alignments)
1931.586 Million cell updates/sec

Title: US-09-674-779b-2

Sequence: 1 MMHIIQIAAAAAAALSVTFM.....IAFGNNEPAMQNRRLASY 172

Scoring table:

BLASUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Command line parameters:
-MODE=frame.p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO_spool/US09674779/unat.06072003.121615.9891/app.query.fasta.1.327
-DB=Published_Applications_NA -OFM=fastap -SUFFIX=p2n.rnpb MIMATCH=0.1
-LOOPEXT=0 -LOOPEXT-0 -UNITS-bits -START=1 -END=1 MATRIX=blasum62
-TRANS=human40.cdi -LIST=100 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR.MIN=0 -ALIGN=40 -MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09674779 -ECGN_1.1.122 -eunat.06072003.121615.9891
-NCPU=6 -ICPU=3 -NO_MAP -LARGEQUERY -NEG.SCORES=0 -NAIT -DSPLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOR=6 -DELEXT=7

Database: Published_Applications_NA:

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
10:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
11:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
12:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:
13:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
14:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	30.7	1830121	9	US-10-329-960-1
2	204.5	23.8	779	10	US-09-747-348-1
3	176.5	20.5	1830121	9	US-10-329-960-1
4	160	18.6	984	9	US-10-216-338A-5

5	149	17.3	678	9	US-10-156-761-3232	Sequence 3322, Ap
6	149	16.8	861	9	US-10-156-761-1	Sequence 1, Appl1
7	145	15.9	730	10	US-10-216-338A-7	Sequence 7, Appl1
8	136.5	15.9	1035	10	US-09-452-559-11	Sequence 11, Appl1
9	136.5	15.7	640681	10	US-09-790-988-1	Sequence 1, Appl1
10	135.5	15.7	561	10	US-09-974-300-1902	Sequence 1902, Ap
11	135.5	13.2	1143	9	US-09-998-278-25	Sequence 25, Appl
12	107	12.4	669	10	US-09-974-300-1889	Sequence 1889, Ap
13	101.5	11.8	717	9	US-10-010-160-11	Sequence 11, Appl
14	93.5	10.9	4268	8	US-08-424-550B-80	Sequence 80, Appl
15	87.5	10.2	8912	8	US-08-424-550B-11	Sequence 11, Appl
16	87.5	10.2	9143	8	US-08-424-550B-390	Sequence 390, Ap
17	87.5	10.2	9143	8	US-08-424-550B-393	Sequence 393, Ap
18	87.5	10.2	8939	10	US-09-742-659-1	Sequence 1, Appl1
19	87.5	10.1	8033	10	US-09-070-927A-121	Sequence 121, Ap
20	87	9.5	2816	10	US-09-886-468-8	Sequence 8, Appl1
21	82	9.5	4224	10	US-09-841-132-486	Sequence 171, Ap
22	82	9.4	2895	10	US-09-841-132-171	Sequence 183, Ap
23	81	9.4	2934	10	US-09-841-132-183	Sequence 58, Appl
24	81	9.3	1086	12	US-10-007-693-58	Sequence 1993, Ap
25	80	9.3	1260	9	US-10-156-761-1993	Sequence 6221, Ap
26	80	9.3	9025608	9	US-10-062-254-323	Sequence 43, Appl
27	80	9.1	315	10	US-09-974-300-6221	Sequence 323, Ap
28	78.5	9.0	1634	12	US-10-007-693-43	Sequence 6208, Ap
29	77.5	9.0	1762	12	US-10-062-254-323	Sequence 9978, Ap
30	77.5	9.0	2631	10	US-10-156-761-1454	Sequence 145, Ap
31	77.5	9.0	1317	10	US-09-815-242-9978	Sequence 145, Ap
32	76.5	8.9	1317	10	US-09-815-242-9978	Sequence 5966, Ap
33	76	8.8	2868	9	US-09-919-039-60	Sequence 4870, Ap
34	76	8.8	10300	10	US-09-960-253-145	Sequence 60, Appl
35	75.5	8.8	465	9	US-10-156-761-5966	Sequence 145, Ap
36	75.5	8.8	2382	10	US-09-815-242-4642	Sequence 4566, Ap
37	75.5	8.8	2409	10	US-09-815-242-8570	Sequence 8570, Ap
38	75	8.7	1335	9	US-10-156-761-1454	Sequence 145, Ap
39	75	8.7	1707	10	US-09-815-242-4085	Sequence 4085, Ap
40	75	8.7	11575	10	US-09-938-956-5	Sequence 2960, Ap
41	74.5	8.7	3990	9	US-10-156-761-2960	Sequence 5966, Ap
42	74	8.6	605	9	US-09-796-692-6743	Sequence 6743, Ap
43	74	8.6	605	9	US-10-040-662-6743	Sequence 13478, A
44	74	8.6	1387	9	US-10-198-846-13428	Sequence 2778, Ap
45	74	8.6	1860	9	US-10-156-761-2778	Sequence 582, Ap
46	74	8.6	2265	9	US-10-037-770-582	Sequence 663, Ap
47	74	8.6	3180	12	US-10-044-090-663	Sequence 2493, Ap
48	74	8.6	3540	9	US-10-128-714-2493	Sequence 7493, Ap
49	74	8.6	3877	9	US-10-128-714-7493	Sequence 6493, Ap
50	74	8.6	5877	9	US-10-128-714-6493	Sequence 493, Ap
51	74	8.6	5877	9	US-10-128-714-493	Sequence 5493, Ap
52	74	8.6	5877	9	US-10-128-714-493	Sequence 79, Appl
53	74	8.6	5877	9	US-10-128-714-493	Sequence 137, Ap
54	74	8.6	38155	9	US-10-114-170-79	Sequence 24222, A
55	74	8.6	48908	9	US-10-114-170-79	Sequence 2679, Ap
56	74	8.5	48908	9	US-10-114-170-137	Sequence 1581, Ap
57	73.5	8.5	1707	9	US-09-918-995-24922	Sequence 6520, Ap
58	73.5	8.5	1707	9	US-09-738-626-2679	Sequence 581, Ap
59	73.5	8.5	1720	9	US-10-128-714-6581	Sequence 5581, Ap
60	73.5	8.5	2787	10	US-09-815-242-6320	Sequence 31, Appl1
61	73.5	8.5	3720	9	US-10-128-714-5581	Sequence 330, Ap
62	73.5	8.5	3720	9	US-10-128-714-5581	Sequence 333, Ap
63	73.5	8.5	6791	9	US-09-771-035A-1	Sequence 1, Appl1
64	73.5	8.5	13856	7	US-08-781-986A-31	Sequence 1889, Ap
65	73.5	8.5	3309400	9	US-09-738-626-1	Sequence 80, Appl
66	73	8.5	4205	9	US-10-037-2270-480	Sequence 480, Ap
67	73	8.5	4226	9	US-10-037-2270-480	Sequence 4513, Ap
68	73	8.5	4473	10	US-09-815-242-4513	Sequence 8064, Ap
69	73	8.5	4509	10	US-09-815-242-8064	Sequence 79, Appl
70	73	8.5	4560	10	US-09-738-626-944	Sequence 4, Appl1
71	73	8.5	5659	12	US-10-002-600-80	Sequence 252, Ap
72	73	8.5	5763	12	US-10-002-600-79	Sequence 56, Appl
73	73	8.5	6002	9	US-10-071-179-4	Sequence 1, Appl1
74	73	8.5	12848	9	US-10-114-170-252	
75	73	8.5	30246	7	US-08-781-986A-56	
76	73	8.5	3309400	9	US-09-738-626-1	
77	72.5	8.4	915	9	US-10-156-761-4240	

```
c 78 72.5 8.4 1674 9 US-10-156-761-3205
79 72.5 8.4 2553 9 US-10-156-761-2383
80 72.5 8.4 10610 9 US-09-994-412-3
81 72.5 8.4 11517 10 US-09-901-106-1
82 72 8.4 885 9 US-10-156-761-1076
83 71.5 8.3 770 9 US-10-198-846-12573
84 71.5 8.3 1323 9 US-10-156-761-6130
85 71.5 8.3 1521 9 US-10-156-761-6926
86 71.5 8.3 7516 10 US-09-917-800A-1664
87 71 8.2 514 9 US-09-918-995-25705
88 71 8.2 546 9 US-10-114-666-9
89 71 8.2 546 10 US-09-895-828-9
90 71 8.2 797 10 US-09-880-107-3806
91 71 8.2 886 9 US-09-996-634-119
92 71 8.2 886 9 US-09-997-181-119
93 71 8.2 886 9 US-09-997-182-119
94 71 8.2 1389 9 US-10-156-761-968
95 71 8.2 1391 10 US-09-925-301-66
96 71 8.2 4056 9 US-10-156-761-2317
97 71 8.2 6694 9 US-10-074-095-700
98 71 8.2 6694 10 US-09-764-860-700
99 71 8.2 7684 10 US-09-070-927A-16
100 71 8.2 8688 9 US-09-938-842A-1748
```

ALIGNMENTS

```
RESULT 1
US-10-329-960-1/c
Sequence 1, Application US/10329960
Publication No. US20030099277A1
```

```
GENERAL INFORMATION:
```

```
APPLICANT: Fleischmann et al.
```

```
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
```

```
TITLE OF INVENTION: Thereof, and Uses Thereof
```

```
FILE REFERENCE: P186P1
```

```
CURRENT APPLICATION NUMBER: US/10/329,960
```

```
PRIOR FILING DATE: 2003-01-02
```

```
PRIOR APPLICATION NUMBER: US 09/643,990
```

```
PRIOR FILING DATE: 2000-08-23
```

```
PRIOR APPLICATION NUMBER: US 08/487,429
```

```
PRIOR FILING DATE: 1995-06-07
```

```
PRIOR APPLICATION NUMBER: US 08/426,787
```

```
PRIOR FILING DATE: 1995-04-21
```

```
NUMBER OF SEQ ID NOS: 1
```

```
SOFTWARE: PatentIn version 3.1
```

```
SEQ ID NO 1
```

```
LENGTH: 1830121
```

```
TYPE: DNA
```

```
ORGANISM: Haemophilus Influenzae
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (4747)..(4747)
```

```
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (9921)..(9921)
```

```
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (10150)..(10150)
```

```
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (36543)..(36543)
```

```
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (36551)..(36551)
```

```
Sequence 3205, Ap
Sequence 2383, Ap
Sequence 3, Appl
Sequence 1, Appl
Sequence 1076, Ap
Sequence 12573, A
Sequence 6130, Ap
Sequence 6926, Ap
Sequence 1664, Ap
Sequence 25705, A
Sequence 9, Appl
Sequence 9, Appl
Sequence 3806, Ap
Sequence 119, Ap
Sequence 119, Ap
Sequence 119, Ap
Sequence 968, Ap
Sequence 66, Appl
Sequence 2317, Ap
Sequence 700, Ap
Sequence 700, Ap
Sequence 16, Appl
Sequence 1748, Ap
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (53369)..(53369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139100)..(139100)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
```

```
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)
```

Alignment Scores:

Pred. No.:	6.02e-23	Length:	1830121
Score:	264.00	Matches:	51
Percent Similarity:	68.93%	Conservative:	20
Best Local Similarity:	49.51%	Mismatches:	32
Query Match:	30.66%	Indels:	0
DB:	9	Gaps:	0

US-09-674-779b-2 (1-172) x US-10-329-960-1 (1-1830121)

```
QY 70 ValtyrPheaspPheaspSeraspGluIlelyspProGlnAlaAlaIleleuaspGlu 89
Db 401840 GTTATTTCGTTTGTGTAATAATATGACATTACTGTGTAATACGTTCAATCTTAGACGG 401781
QY 90 GlnAlaGlnPheleuthrThrAsnGlnThrAlaArgValleuValAlaGlyIsthrasp 109
Db 401780 CAGCTGCATATTTAAATGCAACGCCAGCTGTAAAGTATTAAGTAAAGTAAAGTAAAGTAA 401721
QY 110 GlnArgGlySerArgIleuArgIleuArgIleuArgIleuArgIleuArgIleuArgIleu 129
Db 401720 GAACGTGTACACAGATATCAACATCGCATTTAGGCCACGTCGTGCATGCATGATA 401661
QY 130 AsnTyrIleuGlnGlyIleuGlnGlnAlaSerValGluIleIleSerPheGlyGlu 149
Db 401660 GGTATTATTACCTGTAAAGGTGTGATGCTGTAAATTTAGGCACACTATCTTACGCTGA 401601
QY 150 GluArgProIleAlaPheGlyThrAsnGlnGlnAlaIleIleSerPheGlyGlu 169
Db 401600 GAAACCTGCAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 401541
QY 170 LeuSerTyr 172
Db 401540 TTACGCTGAC 401532
```

RESULT 2

```
US-09-747-348-1
; Sequence 1, Application US/09747348
; Patent No. US20020123067A1
; GENERAL INFORMATION:
; APPLICANT: Andrew D. MURDIN, Raymond P. OOMEN, Joe WANG, Pamela DUNN
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses th
; FILE REFERENCE: 77813-37
; CURRENT APPLICATION NUMBER: US/09/747,348
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 60/171,525
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 779
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(676)
US-09-747-348-1
```

Alignment Scores:

Pred. No.:	1.14e-19	Length:	779
Score:	204.50	Matches:	46
Percent Similarity:	58.82%	Conservative:	24
Best Local Similarity:	38.66%	Mismatches:	46

Query Match: 23.75% Indels: 3
DB: 10 Gaps: 2
US-09-674-779b-2 (1-172) x US-09-747-348-1 (1-779)
QY 54 ASPASNPASGLUHFVALLYSALALASERLYSEUProSerLeu-----ValTYR 71
DB 302 GATTCCAAAGMAAAGAACATACAGTCACAGCCAGTTGCGATTCTGTAATTCACC 361
QY 72 PheASPheASPseraspclulileys---ProGlnAlaAlaAlleleuaspclugln 90
DB 362 TTTCCTACACAGCCTATACATTAAGTGAAGAACCTGCCTTCCTACAGACTTG 421
QY 91 AlaGlnPheLeuThrThrsnglnThraAlaArgValLeuValAlaGlyIsthrspglu 110
DB 422 GTTCACTACATGAAGAAACCCGAAAGCTACACTGATGGAATGACATCTGACGAG 481
QY 111 ArgGlySerArggluThrsnmetSerLeuGlyLArgArgAlaValAlaValArgasn 130
DB 482 CGTGGAGCTGCATCTATACCTTCTTGAAGACGACGACGACCAATCGATTAAAGAG 541
QY 131 TyrlleuLeuGlylysglyleasnGlnAlaSerValglulileSerPhegluglu 150
DB 542 CATCTCCGAAGAGGAAATCTCTGCAATCTGCTATCTACTATTCTACGCGAAAGAA 601
QY 151 ArgProIleAlaPheglYThrsnGlnGluAlaTrpSerGlnasnArgArgAlaGlu 169
DB 602 CATCTTTAAATTCGGGACACAAAGACTAGCATGGACAAATGCGCCCTACAGAG 658
RESULT 3
US-10-329-960-1
Sequence 1, Application US/10329960
Publication No. US20030099277A1
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
FILE REFERENCE: P8186p1
CURRENT APPLICATION NUMBER: US/10/329, 960
CURRENT FILING DATE: 2003-01-02
PRIOR APPLICATION NUMBER: US 09/643, 990
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: US 08/487, 429
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426, 787
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4416)..(4416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

US-09-b/4-779B-2 (1-172) X US-10-329-960-1 (1-1830121),

```

QY      33 ProAnaAlaProThnGlyTyr-----ThrGlyValIleTyr    46
Db      1235275 CCAAAATACCAGCATTAACCTAACACCCCTGGATTGGTTTCATGCATGCGGGGTATTTCTTAC 1235334
QY      47 -----ThrGlyValAlaProLeuValAspAsnAspDluthrValLysAlaLeuAla    63
Db      1235335 CGTTTTGGTCGAAGCGCAGCACCGATGTTGTGCAGCACACCTGMAAGCTTGTAAGCAAACCTTTC 1235394
QY      64 SerLysLeuProSerLeuValTYRPhaSPheASPserASpDluIleLysProGlnAla    83
Db      1235395 AEC---TTAAACTGTGATGTAACCTTCCTGCCATTGTGGTAAAGCAAAACTTAAACCTCAACGA 1235451
QY      84 AlaAlaIleLeuAspSgUGlnAlaGlnPheLeuThrThraSnglOthAlaArgAlaLeu    103
Db      1235452 CAAGCAACAATTAGAACCACATCTATTGGGAAATGTCCAAAGTTAAAGTGCANAAAGTACT 1235511
QY      104 ValAlaGlyHisThrAspGluArvgLySerArvgGluTYRAsmetserLeuGyluarG    123
Db      1235512 GTTGTGTGTACACTGCACCTATTGGTGTGCAGCGGTGCYAACGTAAACCTTCTCAGAA 1235571
QY      124 ArgAlaValAlaValAArgasnTYrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGlu    143
Db      1235572 CGTGCAATTCAGTAGCTAACTACTCTTGTGCTAAAGGTTGTGCAGCAGCAGCAATCTCT 1235631
QY      144 IleIleserPheGlyGlnGluArGProlIealaphGlyThrAsnGlnGlu-----    160
Db      1235632 GCAACTGTGACGGTAAAGCAACCCAGTAAGTGTGCACACTTGTGCACCAAGTTAAAGGT 1235691
QY      161 -----AlarTPSerGlnAsnArGatGalGluLeuSer    171
Db      1235692 CGTAAAGCGCTTATTCGCTGTGTTTGGTCCAGACCGTCGTAAGAAATGGC 1235742

RESULT 4
US-10-216-338A-5
Publication No. US20030113757A1
GENERAL INFORMATION:
APPLICANT: Czajka, John
TITLE OF INVENTION: Rapid and Specific Detection of Campylobacter
FILE REFERENCE: MD108JDSNA
CURRENT APPLICATION NUMBER: us/10/216, 338A
CURRENT FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: 60/310, 882
PRIOR FILING DATE: 2001-08-08
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Microsoft Office 97
SEQ ID NO 5
LENGTH: 984
TYPE: DNA
ORGANISM: Campylobacter coli
```


US-09-674-779B-2 (1-172) x US-09-452-599-11 (1-730)

```

RESULT 9
US-10-169-953-1
; Sequence 1, Application US/10169953
; Publication No. US20030044915A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Thierry BAUSSANT
; APPLICANT: Pascal JEANNIN
; APPLICANT: Yves DELNESTE
; APPLICANT: Francois LAMNY
; APPLICANT: Jean Yves BONNEFOY
;
; TITLE OF INVENTION: METHOD FOR PREPARING A POLYPEPTIDE SOLUBLE IN AN AQUEOUS SOLVENT
; TITLE OF INVENTION: IN THE ABSENCE OF DETERGENT

```

QY 66 leuproserleuValIyrpheaSpaShSerSpIunIleYsProGInAlAla 85
 643 CTGAAGTCAGCTTCTGTTCAATTCACAAAGCTACCCGGAACCGGAAGTCACAG 702
 QY 86 IleleuSpelugInAlaGInpheleuThr-----AsnGInThrAlaYValIeu 104
 703 GCTTGATGACTGCTGTACCTACAGTCAGCAACATGATCCGGAAGAGCTGCCGCTT 762
 Db 104 ValAlaGlyHisThrAspGluArgGlySerArgIuIuYrAsnMetSerIeuGlyGluArg 122

```

1      RESULT 10
2      US-09-790-988-1/C
3      ? Sequence 1, Application US/09790988
4      ? Patent No. US20020127687A1
5      ?
6      ? GENERAL INFORMATION
7      ?
8      ? APPLICANT: SHIGEMOBU, SHUJI
9      ? APPLICANT: WATANABE, HIDEKI
10     ? APPLICANT: HATTORI, MASAHIRO
11     ? APPLICANT: SAKAKI, YOSHIYUKI
12     ?
13     ? TITLE OF INVENTION: GENOME DNA OF BACTERIAL STYMBION OF APHIDS
14     ?
15     ? FILE REFERENCE: 081356/0159
16     ?
17     ? CURRENT APPLICATION NUMBER: US/09/790,988
18     ?
19     ? CURRENT FILING DATE: 2001-02-23
20     ?
21     ? PRIOR APPLICATION NUMBER: JP2000-107160
22     ?
23     ? PRIOR FILING DATE: 2000-04-07
24     ?
25     ? NUMBER OF SEQ ID NOS: 7
26     ?
27     ? SOFTWARE: Patentlin Ver. 2.1
28     ?
29     ? SEQ ID NO 1
30     ?
31     ? LENGTH: 640681
32     ?
33     ? TYPE: DNA
34     ?
35     ? ORGANISM: Buchnera sp.
36     ?
37     ? US-09-790-988-1

```

DB 305180 1A1CCG11A
.....

RESULT 11


```
US-09-974-300-1902
; Sequence 1902, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, ID Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1902
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1902
```

```
Alignment Scores:
Pred. No.: 9.55e-07 Length: 561
Score: 113.50 Matches: 35
Percent Similarity: 44.72% Conservative: 20
Best Local Similarity: 28.46% Mismatches: 63
Query Match: 13.18% Indels: 5
Gaps: 3
```

US-09-674-779B-2 (1-172) x US-09-974-300-1902 (1-561)

```
QY 53 ValAspAsnAspGluThrValIysAlaLeuAlaSerLysLeuProSerLeuValIyrPhe 72
Db 150 GTCATACGAGAGCTACAGACGAGCGCTCCCTTCCTCCATCGAGATATAATCTTTTC 209
QY 73 AspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGluGlnAlaGln 92
Db 210 GATTCGCGAAGAACGGAGATCCGCGACGACGACATTCCTCGCGCAAGAGATATCCGAC 269
QY 93 PheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlnIleThrAspGlu----- 110
Db 270 CTTCGCGATTCGAGACCCGCCCGCAATATCTTAATCGCGGCGCATACGAGAAATGTCGC 329
QY 111 ---ArgGlySerArg---GluTyrAsnMetSerLeuGluArgArgAlaValAlaVal 128
Db 330 ATTCGAATTCATTAATCAATTAATCAATTAATGCGATTAAGCGTGAATGCGGCTCATTTTC 389
QY 129 ArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerVal---GluIleIleSerPhe 147
Db 390 ATGGGGCTTTAATTCGAAATCCAAACCTTGACGCCCAAGATCTTCACGCGGAAAGCGCTAC 449
QY 148 GlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaIrrPserGlnAsnArgArg 167
Db 450 GGGGAATTTAAACCGATCGCTCAATATGACACCGAAGAGAGAGAGAGAGAGAGAGAGAG 509
QY 168 AlaGluLeu 170
Db 510 GTTGAAATTC 518

RESULT 12
US-09-998-279-25
; Sequence 25, Application US/0998279
; Publication No. US20030083287A1
; GENERAL INFORMATION:
; APPLICANT: BURGESS, NICOLA A.
; APPLICANT: GARCIA, MIGUEL M.
; APPLICANT: KIRKE, DAVID F.
; APPLICANT: MEYERS, NICHOLAS L.
; APPLICANT: WILLIAMS, PAUL
; TITLE OF INVENTION: gins
; FILE REFERENCE: GMS0081
```

```
;; CURRENT APPLICATION NUMBER: US/09/998,279
;; PRIOR FILING DATE: 2001-11-30
;; PRIOR APPLICATION NUMBER: 60/250,288
;; PRIOR FILING DATE: 2000-11-30
;; NUMBER OF SEQ ID NOS: 26
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 25
;; LENGTH: 1143
;; TYPE: DNA
;; ORGANISM: Porphyromonas gingivalis
US-09-998-279-25
```

```
Alignment Scores:
Pred. No.: 2.55e-05 Length: 1143
Score: 107.00 Matches: 40
Percent Similarity: 42.75% Conservative: 16
Best Local Similarity: 30.53% Mismatches: 49
Query Match: 12.43% Indels: 26
Gaps: 6
```

US-09-674-779B-2 (1-172) x US-09-998-279-25 (1-1143)

```
QY 35 ProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGlyValAlaProLeuValAsp 54
Db 799 CCTACACAGCCTTACA---GTTACTCGTGTAGTC-----GTTGAC 834
QY 55 AsnAspGluThrValIysAlaLeuAlaSerLysLeuProSerLeuValIyrPheAspPhe 74
Db 835 AAT-----GTGGTTACTTCGCTATC 855
QY 75 AspSerAspGluIleLysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeu 94
Db 856 AATGTCGAAGATGATGCTTAATCAGAAATCAATGTTCAATACACTGCAATATGCG 915
QY 95 ThrThrAsnGlnThrAlaArgValLeuValAlaGlnIleThrAspGluArg---GlySer 113
Db 916 AAGACCAAC---AACGACCGATCAAGGTAGTGTACGCTGACGAGAGAGAGAGAGAGAG 972
QY 114 ArgGluTyrAsnMetSerLeuGluArgArgAlaValAlaValArgAsnTyrLeuLeu 133
Db 973 GCGGCGTATTAACATGTAAGCTTTCAGAGCGCTGCGCAAGAGCGGTACCAAGATGTTGA 1032
QY 134 GlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluArgProIle 153
Db 1033 AAGTATGATGTTCTCTCGGATCGCATTAATGAATGAAGGCGTCAAGAGCAATTC 1092
QY 154 AlaPheGlyThrAsnGluGluAlaIrrPserGln 164
Db 1093 -----TATGAAGAGAACGCTTGAAATCG 1116

RESULT 13
US-09-974-300-1889
; Sequence 1889, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, ID Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1889
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1889
```

ATTORNEY/AGENT INFORMATION:
NAME: POBEMSKI, PRISCILLA E.

ATTORNEY/AGENT INFORMATION:
NAME: POREMBSKI, PRISCILLA E.

SEQ ID NO 486
 LENGTH: 4224
 TYPE: DNA
 ORGANISM: Chlamydia pneumoniae
 us-09-841-132-486

Alignment Scores:

Pred. No.: 0.819 Length: 4224
 Score: 82.00 Matches: 42
 Percent Similarity: 34.29% Conservative: 18
 Best Local Similarity: 24.00% Mismatches: 81
 Query Match: 9.52% Indels: 34
 Gaps: 7

US-09-674-779b-2 (1-172) x US-09-841-132-486 (1-4224)

QY 15 SerValLeuThrPheMet-----ThrGlyCysAlaAsnLys 26
 Db 1276 GCATCTCAGCCCTTTATGATTCAGAGCCGTAAGCATAAAGAGAGATTCGATTCGCA 1335
 QY 27 SerThrSerGlnMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyr 46
 Db 1336 AACACCAAGAAAGTCAAGCTCCTGATATGCTGACAGACGTAAGTGCGTGCATCTAT 1395
 QY 47 ThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeu 66
 Db 1396 GCTACCAATGTACTCTTACTGGAACGCTCC----- 1428
 QY 67 ProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIle 86
 Db 1429 -----CTGACCTTTCAGCGCAATCTGCTGCACTTACAGAGGGCGATC 1473
 QY 87 LeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGly 106
 Db 1474 TATACAGAAATCGAAGATTCTTACTCTACAGGAAGTACGAGAACCGGACCTTCAGACA 1533
 QY 107 HisThrAspGluArgGlySerArgGluTyr-----AsnMetSerLeu---GlyGlu 122
 Db 1534 AATACAGCAAGACAGCGCGCCCTTATATCTTAAAGCAACACTCTGCTGCTGTAAT 1593
 QY 123 ArgGlnAlaValAlaValAlaArgAsnTyrLeuGlyLysGlyIleAsnGlnAlaSerVal 142
 Db 1594 ACCAACCTGCTCTTTCAGGAAACAAAGCTACGGCGGAGTATCTTCAGCAAAATCAA 1653
 QY 143 Glu-----IleIleSerPheGlyGluGlnArgProIleAlaPheGlyThr 157
 Db 1654 GAGGGTTCGGTGGGCGCAATCTATCGTTCTTGAAGTACGACATCTGTAAGT-----ACT 1707
 QY 158 AsnGluGlnAlaTyr---SerGlnAsnArgArgAlaGluLeuSer 171
 Db 1708 AAAAAAGAGCTCGATGATGAAGATAAGAAACGTAAGTCTCTCT 1752

RESULT 23

US-09-841-132-171
 Sequence 171, Application US/09841132
 Patent No. US20020061848A1
 GENERAL INFORMATION:
 APPLICANT: Bhatia, Ajay
 APPLICANT: Skelky, Yasir A.W.
 APPLICANT: Probst, Peter
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
 TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
 FILE REFERENCE: 210121.469C8
 CURRENT APPLICATION NUMBER: US/09/841.132
 CURRENT FILING DATE: 2001-04-23
 NUMBER OF SEQ ID NOS: 599
 SOFTWARE: FastSeq for Windows Version 3.0/4.0
 SEQ ID NO 171
 LENGTH: 2895
 TYPE: DNA
 ORGANISM: Chlamydia
 us-09-841-132-171

Alignment Scores:

Pred. No.: 0.63 Length: 2895
 Score: 81.00 Matches: 39
 Percent Similarity: 37.58% Conservative: 23
 Best Local Similarity: 23.64% Mismatches: 75
 Query Match: 9.41% Indels: 28
 Gaps: 7

US-09-674-779b-2 (1-172) x US-09-841-132-171 (1-2895)

QY 8 AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLysSer 27
 Db 829 GCTTCTGATGAGAGAGCAATTAAGTAACTACTCCCTAGATGTTACAGCAATCGT--- 885
 QY 28 ThrSerGlnMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThr 47
 Db 886 ---GTAAGATCTTTTGTGACATATACAAAAATTAAGGCGGAGCTATTACGCT 942
 QY 48 GlyValAlaProLeuValAspAsnAspGluThr-----ValLysAlaLeuAlaSerLys 65
 Db 943 CCTGTAGTACCTCTAGTGATGAGCCCTACCTCTTATTAACATATGCCAATAT 1002
 QY 66 LeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAla 85
 Db 1003 AAGGGGGCGCTATCTATATATACGGAACCGATAAC-----TCCAAAT 1047
 QY 86 IleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAla 105
 Db 1048 TCTGCCGACCGCCATGATATTTTAATGAATAATTTGACATATGTAATATGCA 1107
 QY 106 GlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAla 125
 Db 1108 AATGTACACAGTACGCTACGTAATCTCTCT-----AGAGAAATGCA 1149
 QY 126 ValAlaValAlaArgAsn-----TyrLeuGlyLysGlyIleAsnGln----- 139
 Db 1150 ATACAGTAGCAAGCTCTCTGTAATTTCTATTAAGACAGGAGTACCAAAATTTA 1209
 QY 140 -----AlaSerValGluIleIleSerPheGlyGluGlnArgProIleAlaPheGly 156
 Db 1210 ATTTTATGATTCCTATTGAAGATTAGCAATGCAAGG-----GTCTGTGCTCC 1257
 QY 157 ThrAsnGluGlnAla 161
 Db 1258 TTCATATAGGAAGCT 1272

RESULT 24

US-09-841-132-183
 Sequence 183, Application US/09841132
 Patent No. US20020061848A1
 GENERAL INFORMATION:
 APPLICANT: Bhatia, Ajay
 APPLICANT: Skelky, Yasir A.W.
 APPLICANT: Probst, Peter
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
 TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
 FILE REFERENCE: 210121.469C8
 CURRENT APPLICATION NUMBER: US/09/841.132
 CURRENT FILING DATE: 2001-04-23
 NUMBER OF SEQ ID NOS: 599
 SOFTWARE: FastSeq for Windows Version 3.0/4.0
 SEQ ID NO 183
 LENGTH: 2934
 TYPE: DNA
 ORGANISM: Chlamydia
 us-09-841-132-183

Alignment Scores:

Pred. No.: 0.643 Length: 2934
 Score: 81.00 Matches: 39
 Percent Similarity: 37.58% Conservative: 23
 Best Local Similarity: 23.64% Mismatches: 75
 Query Match: 9.41% Indels: 28

[illegible]

```

OY      32 MetValAlaProAsnAla---ProThGly-----TyrThcGlyValIleTyr 46
      ::::|||||
Db      532 ATTCGACGTCGAAGCAGCTCTTCACACGACAGAAAAAAGATAGTACAGCATTTCCTT 591
OY      47 ThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeu 66
      :|||
Db      592 ATTCGGAACGCCCTTTAAAGAGTCTCTAAG-----AAGTTCCTCCATAAACA 645
OY      67 ProSerValTyrPheAspPheSerSerPgluIleLysProGlnAlaAlaIle 86
      ::::
Db      646 GGCACAGCTTATC-----GATTGGATACCAAGAAATTGTAGGCCAA----- 687
OY      87 LeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaIaArgValLeuValAlaGly 106
      :687-----
Db      687 ----- 687
OY      107 HisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaVal 126
      ::::|||||
Db      688 -----CATCAGGGAGCTCCTACTATAT-----ACTATAGGCGACGCCGACAGACTT 732
OY      127 AlaValArgAsn-----TyrLeuLeuGlyLysGlyLysGlnGlnAlaSer 141
      ::::
Db      733 GATCTTCGACGATCCGAGAAACCTGTATATGTTGGGAAAAATATATAGAGGAAAAATAGC 792
OY      142 ValGluIleIleSerPheGlyGluGluArgPro 152
      ::::|||||
Db      793 ATTATATTTGTGAGG---GGGGAGAGCAATCCC 822

RESULT 26
US-10-156-761-1993
? Sequence 1993, Application US/10156761
? Publication No. US20030119018A1
? GENERAL INFORMATION:
? APPLICANT: OMDRA, SATOSHI
? APPLICANT: IKEDA, HARUO
? APPLICANT: ISHIKAWA, JUN
? APPLICANT: HORIKAWA, HIROSHI
? APPLICANT: SHIBA, TADAYOSHI
? APPLICANT: SAKAKI, YOSHIYUKI
? APPLICANT: HATTORI, MASAHIRO
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-262
? CURRENT APPLICATION NUMBER: US/10/156,761
? PRIOR FILING DATE: 2002-05-29
? PRIOR APPLICATION NUMBER: JP 2001-204089
? PRIOR FILING DATE: 2001-05-30
? PRIOR APPLICATION NUMBER: JP 2001-272697
? PRIOR FILING DATE: 2001-08-02
? NUMBER OF SEQ ID NOS: 15109
? SEQ ID NO 1993
? LENGTH: 1260
? TYPE: DNA
? ORGANISM: Streptomyces avermitilis
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1)..(1260)
US-10-156-761-1993

Alignment Scores:
Pred. No.: 0.236 Length: 1260
Score: 80.00 Matches: 43
Percent Similarity: 41.30% Conservative: 33
Best Local Similarity: 23.37% Mismatches: 84
Query Match: 9.29% Indels: 24
DB: Gaps: 6

US-09-674-779B-2 (1-172) x US-10-156-761-1993 (1-1260)

```



```

Db 574 AAGGTGACCAAGCAGCTGATGCTGATGCGGAGCG-----ACCGCGGACAG 621
Qy 46 TTTTThrglValAlaProLeuValAspAsnAspGluThrValLysAlaLeu-----62
Db 622 TACACCTCGGTGGGACCAAGAGAGACCTGACCGACAGGCTCAACAGCTGGTGGACCGC 681
Qy 63 ---AlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLysPro 81
Db 682 GCCCGCCGACAAAGGTGGTGGACCGCGGTCCCGTGGACGGCGCGCGCGCGGACGAGCGG 741
Qy 82 GlnAla-----AlaAlaIleLeuAspGluGlnAlaGlnPhe-----93
Db 742 CCGACGCTGAAGTCCGGCTCTTCACGAGCGCGGAGTGTGCCACGACCGCTGTATC 801
Qy 94 -----LeuThrThrasnGlnThrAlaArgValLeuValAlaGlnHisThrAspGlu 110
Db 802 CGCGTGGACGTGAAGCGCGGCGGACGAGCTCGCGCTCGGTGAGCTCTCCGCGACCGT 861
Qy 111 ArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsn 130
Db 862 CAGGTGAACCCGAGCTACGCGGGGTGTGCTG-----CGGCGGTGACCGTGCACAC 912
Qy 131 -----TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGly 148
Db 913 CGTGAAGTCTCGCGGTGAGGCGCGGCGGACGCGCGCTGAGCGAGTGTGTGCGAGCGGT 972
Qy 149 GluGluArgProIleAlaPheGlyThrAsnGluGluAlaTyrSerGlnAsnArgArgAla 168
Db 973 CTGGCTATTCGAAAGCCGAGAGACGACGACGAGCGACCGCGGACCGGACCGGTGTCTC 1032
Qy 169 GluLeuSerTyr 172
Db 1033 CAGGTCTCCAC 1044

RESULT 27
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication NO: US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

Alignment Scores:

```

Pred. No.: 2.66e+05
Score: 80.00
Percent Similarity: 41.308
Best Local Similarity: 23.378
Query Match: 9.29%
Gaps: 6
Length: 9025608
Matches: 43
Conservative: 33
Mismatch: 84
Indels: 24

```

```

US-09-674-779b-2 (1-172) x US-10-156-761-1 (1-9025608)
Qy 6 GlnIleAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsn 25
Db 2445562 GAGATCCGCGCCCAAGGGCGTCCGGCTCCACATCGACACCTCGCTGTGCGGACAGC...
Qy 26 LysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIle 45
Db 2445622 AAGCTGACGACAGAGGCTGAGCTGATCGCGGAGCG-----ACGCGCGGACG 2445669
Qy 46 TTTTThrglValAlaProLeuValAspAsnAspGluThrValLysAlaLeu-----62
Db 2445670 TACACCTCGGTGGACCAAGAGAGCTTCACCGGACCGTCAACAGCTGGTGGACCGC 2445729
Qy 63 ---AlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLysPro 81
Db 2445730 GCCCGCCGACAAAGTGTGGACGCGCGGTGGACGGCGCGCGCGCGGACGAGCGG 2445789
Qy 82 GlnAla-----AlaAlaIleLeuAspGluGlnAlaGlnPhe-----93
Db 2445790 CCGACGCTGAAGTCCGGCTCTTCACGAGCGCGGAGTGTGCCACGACCGCTGTATC 2445849
Qy 94 -----LeuThrThrasnGlnThrAlaArgValLeuValAlaGlnHisThrAspGlu 110
Db 2445850 CGCGTGGACGTGAAGCGCGGCGGACGAGCTCGCGCTCGGTGAGCTCTCCGCGACCGT 2445909
Qy 111 ArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsn 130
Db 2445910 CAGGTGAACCCGAGCTACGCGGGGTGTGCTG-----CGGCGGTGACCGTGCACAC 2445960
Qy 131 -----TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGly 148
Db 2445961 CGTGAAGTCTCGCGGTGAGGCGCGGCGGACGCGCGCTGAGCGAGTGTGTGCGAGCGGT 2446020
Qy 149 GluGluArgProIleAlaPheGlyThrAsnGluGluAlaTyrSerGlnAsnArgArgAla 168
Db 2446021 CTGGCTATTCGAAAGCCGAGAGACGACGACGAGCGACCGCGGACCGGACCGGTGTCTC 2446080
Qy 169 GluLeuSerTyr 172
Db 2446081 CAGGTCTCCAC 2446092

RESULT 28
US-09-974-300-6221
; Sequence 6221, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Beika, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 2000-10-06
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6221
; LENGTH: 315
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6221

```

Alignment Scores:

```

Pred. No.: 0.0437
Score: 78.50
Percent Similarity: 44.448
Best Local Similarity: 25.568
Query Match: 9.128
Length: 315
Matches: 23
Conservative: 17
Mismatch: 45
Indels: 5

```

```

Oy      8 AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyGlySAlaLeuLysSer 27
Db      390 GCTTCCTCATGAGGAGACCAATTAAAGTAACACTGCGCCATGATGTTAAGGCATGCT---- 334
Oy      28 ThrSerGlnValMetValAlaProSAlaAlaProThrGlyTyrThrGlyValIleTyrThr 47
Db      333 ---GGTAGACATCTTTTATTACTGACATATACAAAAAATTTATGGCGAGCTATTATTCGCT 277
Oy      48 GlyValAlaProLeuValaAspAnaSpGluThr-----ValLysAlaLeuAlaSerLys 65
Db      276 CCTGAGTATTCACCTAGGAGATATATGGCCCTACTACTCTATTATATTAACAATATATGCCAATAAT 217
Oy      66 LeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaAla 85
Db      216 AAGGGGGGCGCTATCTATATATAGACGGAACACAGTAC-----TCCAAATTT 172
Oy      86 IleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValIleValAla 105
Db      171 TCTGCCGACGCCCATGCTATATATTTTATGAAAATATGTGCACTAATGTACTAATTAATGCA 112

```

Alignment Scores:	
Pred. No.:	0.922
Score:	1762
Percent Similarity:	77.50
Best Local Similarity:	Matches: 46
Query Match:	Conservative: 22
DB:	Mismatches: 81
	Indels: 32
	Gaps: 5
	Gaps: 12

US-09-674-779B-2 (1-172) x US-10-062-254-323 (1-1762)

QY 4 HisIleGlnIleAlaAlaAlaAla-----AlaSerValLeuThrPheMet 20
||||| ||||| ::|||:: ||||| |||||:: |||
Db 523 CACATCGCATGGCCTTATCGGCTTCAAACTTCTCGGCTTGAGGAACTGACGATCATG 466

QY	21	ThrGlyCysAlaAsnLysSerThSerGlnValMetValAlaProAsnAlaProThGly	40
Db	463	TGGTGTCTGCTGCTGCTGGG-CAGAGACGGCGCGCGCGAGAGCGGTCCCGAAGGA	405
QY	41	TyrThrGlyValIleTyrThGlyValAlaProLeuValAlaAspAsnAP-----	56
Db	404	GCGCAGCGCGGCTCTCTCTCGGCGTCCGGACGTCCTCGATGGAGATGGCCCGCGCG	345
QY	57	---GluThrValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAsp	75
Db	344	GAGCGGAGATCGACGGCGCGCTCCCGCTCTTACGGCGCG-----	300
QY	76	SerAspGluLysProGlnAlaAlaAlaLeuAspGluGln-----AlaGlnPhe	93
Db	299	---GATGACCTTCGCGCGCGACGCGAGCTTACGCGCGCGCTCGCTCCGAGT	243
QY	94	LeuThrThrAsnGlnThrAlaArgValLeuValAlaGluHisThrAspGluArgGlySer	113
Db	242	GCCGCTCACAGCAGCATATCTCGATGCTCTCGCGGTGCGCTACCGACATCTTCGCGCGCG	183
QY	114	ArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValAlaArg-AsnTyrLeu	133
Db	182	CGATTCGCGAGACGTCGCGCTCTCGACCGAGACCTCGACCGCGCCGACCTGCGATTTTGT	123
QY	133	ugLYysgLYleAsnGlnAlaSerValGluIleIleSerPheGlyGluArgProIle	153
Db	122	TGGGGGTTTGGATGGAGGAAGCATCATCAT-----	92
QY	153	eAlaPheGlyThrAsnGluAlaTrpSerGlnAsnArgAlaGluLeuSerTyr	172
Db	91	-----CAGAGAGAACCTTGTAAGTACAGGCGCGAGCGCAGAGCTCGGAT	47
RESULT 31			
US-10-156-761-6208			
Sequence 6208, Application US/10156761			
Publication No. US20030119018A1			
GENERAL INFORMATION:			
APPLICANT: OMKRA, SATOSHI			
APPLICANT: IKEDA, HAROO			
APPLICANT: ISHIKAWA, JUN			
APPLICANT: HORIKAWA, HIROSHI			
APPLICANT: SHIBA, TADAYOSHI			
APPLICANT: SAKAKI, YOSHIYUKI			
APPLICANT: HATTORI, MASAHIRA			
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES			
FILE REFERENCE: 249-262			
CURRENT APPLICATION NUMBER: US/10/156,761			
CURRENT FILING DATE: 2002-05-29			
PRIOR APPLICATION NUMBER: JP 2001-204089			
PRIOR FILING DATE: 2001-05-30			
PRIOR APPLICATION NUMBER: JP 2001-272697			
PRIOR FILING DATE: 2001-08-02			
NUMBER OF SEQ ID NOS: 15109			
SEQ ID NO 6208			
LENGTH: 2631			
TYPE: DNA			
ORGANISM: Streptomyces avermitilis			
FEATURE:			
NAME/KEY: CDS			
LOCATION: (1)..(2631)			
US-10-156-761-6208			
Alignment Scores:			
Prod. No.:	1,73	Length:	2631
Score:	77.50	Matches:	53
Percent Similarity:	34.55%	Conservative:	23
Best Local Similarity:	24.09%	Mismatches:	87
Query Match:	9.00%	Indels:	57
Db:	9	Gaps:	8

QY	110	AlAlAlAlAlAlAlLeuSerValLeuThrPheMetThrGlyCysAlaAsnIysSerThrSer	29
		:::	
Db	802	GGCGCGGGGGGGGCGAGAGACGGCGGTCTGTCTCCGCGTGGGGGGCGCGGTGTCTC	861
QY	30	GlnValMetValAlaProAsnAlaProThr-----GlyTrpThrGlyValIleTrpThr	47
		:::::	
Db	862	GGCAGCTCCGGTGGCGGAGCTCTCCGTGACAGCGTGCCTGGGTGGCGGGTCTGGTGGACTGC	921
QY	48	GlyValAla-----ProLeuValAspAsnAspIleThrValIleAlaLeuAlaSer---	64
		::: :::::	
Db	922	GGACTGGCGCGGAGCCACGTGTGATTCACGCGGGGGCGTACGCCCTGACGACCGTA	981
QY	65	-----LysLeuProSer	68
		-----:::	
Db	982	CGGGCTTCGACGGCGGGGGCGGCGACGCGGGCGGGCGGGCGGCGCGAGCGCGGG	104
QY	69	LeuValIyrPhe-----AspPheAspSerAsp	77
Db	1042	ACGCTGACCGCTCTGGGGGAGAGGAGGACGCCCGCTGTGCCCGCTTCCCTCCCC	110
QY	78	GluIleIysPro-----GlnAlaAlaAlaIleLeuAspGlu	89
Db	1102	GAGTCAAGTGGGCCGACCTGACGCGCTTGCCTTGACAGCGCGTGGGGGCAATCGG	116
QY	90	GlnAlaGlnPheLeuThr-----ThrAsnGlnIleAlaArg	101
		:::	
Db	1162	GAGCGCTCCGGGTCCGCTCGTGGACGCGCCCGCGGGGGCGATGGCGGCGCGG	1221
QY	102	ValLeuVal-----AlaGlyHisThrAspGluArgGlySer	113
		:::	
Db	1222	TCCGTTCTCCCGCAGATCGGTGGTGGATTCGCGGGGGCGGCGCACAGGAGAGGGGTG	1281
QY	114	ArgGluTyrAsnMetSerLeuGlyAlaArgAlaValAlaValAlaArgAsnIleLeu	133
		::: :::	
Db	1282	CGCATGTCCCGGTCCGCTCGCTGACCCGCGACTGGCGGGGCTCTCTGACCGCGCGCG	1341
QY	134	GlyIysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgProIle	153
		::::: : :::	
Db	1342	GAGGTGCGCGCGGACCGGCGCGCGGAGTGTGGCGGTGTGACGAGAGACCGCGCGG	1401
QY	154	AlaPheGlyThrAsn-----GluGluAlaIleTrpSerGlnAsnArgAlaGluLeuSerTrp	172
		::: :::	
Db	1402	GAGTACGGCGACGACCTGCGCGGCGGCTGGCGGCTGCCGCGGTGGGGCGACGCGTAC	1461
RESULT 32			
	US-09-815-242-9978	/ Sequence 9978, Application US/09815242	
		/ Patent No. US20020061569A1	
		/ GENERAL INFORMATION:	
		/ APPLICANT: Haselbeck, Robert	
		/ APPLICANT: Ohlsen, Karl U.	
		/ APPLICANT: Zyskind, Judith W.	
		/ APPLICANT: Wall, Daniel	
		/ APPLICANT: Trawick, John D.	
		/ APPLICANT: Carr, Grant J.	
		/ APPLICANT: Yamamoto, Robert T.	
		/ TITLE OF INVENTION: Identification of Essential Genes in	
		/ TITLE OF INVENTION: Prokaryotes	
		/ FILE REFERENCE: EITTRA.011A	
		/ CURRENT APPLICATION NUMBER: US/09/815,242	
		/ PRIOR FILING DATE: 2001-03-21	
		/ PRIOR APPLICATION NUMBER: 60/191,078	
		/ PRIOR FILING DATE: 2000-03-21	
		/ PRIOR APPLICATION NUMBER: 60/206,848	
		/ PRIOR FILING DATE: 2000-05-23	
		/ PRIOR APPLICATION NUMBER: 60/207,727	
		/ PRIOR FILING DATE: 2000-05-26	
		/ PRIOR APPLICATION NUMBER: 60/242,578	
		/ PRIOR FILING DATE: 2000-10-23	
		/ PRIOR APPLICATION NUMBER: 60/253,625	
		/ PRIOR FILING DATE: 2000-11-27	
		/ PRIOR APPLICATION NUMBER: 60/257,931	

Db 37 GTGGAAGGTGTCGCTACGGGCTCCCGAGCTGACGTGGCTGAATTGGAGGGGGCAG 96
Qy 79 ILeYsProGlnAlaAlaAlaLeuAspGluGlnAlaGlnPheLeu----- 94
Db 97 CTGGAGCCTCAGCGCGCAGCGCTCTAGAAATGCTGAGCCGATTCAGGATTCAGCAAT 156
Qy 95 -----ThrThraGlnGlnThraAlaArgValLeuVal 104
Db 157 GTTGTTCATGATATATCAGAGATGATGACACTGATCAGAAATATATGAGGCTCCCTA 216
Qy 105 AlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSer-----Leu 120
Db 217 GACCTGAAATTCACCAAGAAATTCGACATGGAATTAATTAATCAACAAGAAATGTT 276
Qy 121 GlyIuArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
Db 277 CAGGAGCGCGCTGCTTATGCA-----GAGCAATG 306
Qy 141 SerValGluIleIleSerPheGlyGluIuArgProIleAlaPheGlyThrAsnGlu 160
Db 307 GTGGGAGCTAAAGATATTATAGACAGAAAGATGTTCAACTGACAGAAAGATGAA 366
Qy 161 AlaTrpSerGlnAsnArgArgAla 168
Db 367 GCTCTACAGAGAGAGAGAAAGCT 390
RESULT 35
US-10-156-761-5966
; Sequence 5966, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5966
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(465)
US-10-156-761-5966
Alignment Scores:
Pred. No.: 0.219 Length: 465
Score: 75.50 Matches: 38
Percent Similarity: 38.65% Conservative: 25
Best Local Similarity: 23.31% Mismatches: 55
Query Match: 8.77% Indels: 45
Gaps: 7
US-09-674-779b-2 (1-172) x US-10-156-761-5966 (1-465)
Qy 7 ILeAlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLys 26
Db 25 GTGGCGGAGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 75
Qy 27 SerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyr 46
Db 76 AGCGGCTGCGAGAGAGCGCGCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 126

Qy 47 ThrGlyValAlaProLeuValAspAsnAspGluThrValIleAlaLeuAlaSerLysLeu 66
Db 127 -----AccCTGAGAGACTGGCGCGCAGGTG 153
Qy 67 ProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIle 86
Db 154 -----AAGTGCAGGCCGACATCCAGAGGAG 180
Qy 87 LeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThraAlaArgValLeuAlaGly 106
Db 181 GCCGACGAGATCCGCGAGCGGCTGTCAGAGACCTCCAGAAATTCATCCCTCCGACC 240
Qy 107 HisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyIuArgArgAlaVal 126
Db 241 TTGCCACCGACCGCGCGCGCGCGCGAGTGG-----CTCAGACAGGCCAGAGACTAC 291
Qy 127 AlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleSer 146
Db 292 GGC---GGTTCACCTGCTGCGC-----GCCAAGTGGTGGCG 327
Qy 147 PheGlyGluIuArgProIleAlaPhe-----GlyThrAsnGluGlu 160
Db 328 GTGGGAGCAGAGAGGTGTCAAGACGTGGCGGACAGCTCGCGGAGCATGTGAGAG 387
Qy 161 AlaTrpSer 163
Db 388 GGAAGTTCG 396
RESULT 36
US-09-815-242-4642
; Sequence 4642, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.01A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4642
; LENGTH: 2382
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4642
Alignment Scores:
Pred. No.: 2.88 Length: 2382
Score: 75.50 Matches: 38
Percent Similarity: 38.75% Conservative: 24

Best Local Similarity:	23.75%	Mismatches:	55
Query Match:	8.77%	Indels:	33
DB:	10	Gaps:	6

QY		9	AAlAalAAAlAAlAAlLeSeValLeuthrPhemeThrfcLyysAla-----Asnlys	26
Dd		1315	GCACCTGTTGGCAGATATTTCGGTCTCGTCAATGTGCATATAGGACTTGCT	1374
QY		27	SerThSergInValMeTVAlAlaProbsn-----AlAprOthrclYlYr-----	41
Dd		1375	ACACCAACTTCATTATATGTAGGACTGTCGCCGTCTGAANAATGTAATTTATTAAA	1434
QY		42	-----Thrgly	43
Dd		1435	GCGGCGAGATTGTTGAAACGCACACATCAATTCATTCATCGTTTTAGATAAGCGGT	1494
QY		44	VallIeTyThrGlYValAlaProLeuValasp-----AsnspgluThrVallys	60
Dd		1495	ACCATTTCAAAGTGGTCGTCAGTGCAGTCAGCATTCATGTCATCATCAACGCTACA	1554
QY		61	AlAlaLeSerLyLeuProSerLeuValTYrPheasphespeSerpgluLeys	80
Dd		1555	CTRACTTGCTACTGCT-----GAAAAAATTCGTGACAC	1587
QY		81	ProGlnAlaAlaAlaIleLeuasp-----GluGlnAlaGlnPheLeuThrAsngln	98
Dd		1588	CCATTGGCGAGAAGCCATGTGCAATTCAGCAAAAGCAATTAACATTAACTGAGACA	1647
QY		99	ThAlaArgValLeuValAlAGLYHsthrAspgluArgLyserrggluIyrrAsnmet	118
Dd		1648	ACACACTTTTAAAGCAGTACCTGGTCATGTGAATTGAACAACGATTGATCATCACTATA	1707
QY		119	SerLeuglyGluArgAlaValAlaValaIArgsnTrfLeuLeuglysclylleasn	138
Dd		1708	TTCGTTGTGAACCGTAATTAATGCTGACATGATGATTGCTTGCTTAACGATATTTCT	1767
RESULT 37				
US-09-815-242-8570				
Sequence 8570, Application US/09815242				
Patent No. US20020061569A1				
GENERAL INFORMATION:				
APPLICANT: Haselbeck, Robert				
APPLICANT: Ohlsen, Karl L.				
APPLICANT: Zyskind, Judith W.				
APPLICANT: Wall, Daniel				
APPLICANT: Trawick, John D.				
APPLICANT: Carr, Grant J.				
APPLICANT: Yamamoto, Robert T.				
APPLICANT: Xu, H. Howard				
TITLE OF INVENTION: Identification of Essential Genes in				
FILE REFERENCE: ELITRA.011A				
CURRENT APPLICATION NUMBER: Prokaryotes				
PRIOR FILING DATE: 2001-03-21				
PRIOR APPLICATION NUMBER: US/09/815,242				
PRIOR FILING DATE: 2000-03-21				
PRIOR APPLICATION NUMBER: 60/191,078				
PRIOR FILING DATE: 2000-03-21				
PRIOR APPLICATION NUMBER: 60/206,848				
PRIOR FILING DATE: 2000-05-23				
PRIOR APPLICATION NUMBER: 60/207,727				
PRIOR FILING DATE: 2000-05-26				
PRIOR APPLICATION NUMBER: 60/242,578				
PRIOR FILING DATE: 2000-10-23				
PRIOR APPLICATION NUMBER: 60/253,625				
PRIOR FILING DATE: 2000-11-27				
PRIOR APPLICATION NUMBER: 60/257,931				
PRIOR FILING DATE: 2000-12-22				
PRIOR APPLICATION NUMBER: 60/269,308				
PRIOR FILING DATE: 2001-02-16				
NUMBER OF SEQ ID NOS: 14110				
SOFTWARE: FastSeq for Windows Version 4.0				
SEQ ID NO 8570				

```

:      LENGTH: 2409
:      TYPE: DNA
:      ORGANISM: Staphylococcus aureus
:      FEATURE:
:      NAME/KEY: CDS
:      LOCATION: (1)...(2409)
:      US-09-815-242-8570

```

[illegible]

QY 149 lucluarproleialaphelythr-asngluclualatpsergluasnargala 168
 Db 758 AGAGCTAGATTAC---CTGGGACCCCGCCACGAGCGCTGTGCTTCCCGCAGAAAGT 814

Db 9678 TTC 9676
 Search completed: July 6, 2003, 15:04:07
 Job time : 3796 secs

RESULT 40
 US-09-938-956-5/c

; Sequence 5. Application US/09938956
 ; Patent No. US20020142408A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Many, Sigun
 ; APPLICANT: Dicosimo, Deana J.
 ; APPLICANT: Kofas, Matheos
 ; APPLICANT: Odom, J. Martin
 ; TITLE OF INVENTION: Production of Monoterpene
 ; FILE REFERENCE: C11809 US NA
 ; CURRENT APPLICATION NUMBER: US/09/938, 956
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: 60/229, 907
 ; PRIOR FILING DATE: 2000-09-0
 ; PRIOR APPLICATION NUMBER: 60/229, 858
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 5
 ; LENGTH: 11575
 ; TYPE: DNA
 ; ORGANISM: Plasmid
 US-09-938-956-5

Alignment Scores:
 Pred. No.: 41.2 Length: 11575
 Score: 75.00 Matches: 39
 Percent Similarity: 37.89% Conservative: 22
 Best Local Similarity: 24.22% Mismatches: 64
 Query Match: 8.71% Indels: 36
 DB: 10 Gaps: 6

US-09-674-779b-2 (1-172) x US-09-938-956-5 (1-11575)

QY 10 AlaAlaAlaAlaSerValIleuThrPheMetThrGlyCysAlaAsnLysSerThrSer 29
 Db 10095 GCCGGGCGAGCTTCGATCAACGCGCTCAATTCCTGACGGGCTGT----- 10054
 QY 30 GluValMetValAlaProAsnAlaProThrGlyThrGlyValIle--TyrThrGly 48
 Db 10053 -----TTCCTTTGCGGAGTCGCAAAAGCGAGCCCGCCCTTACGCCGGAGGCT 10000
 QY 49 ValAlaProLeuValAsp-----AsnAspGluThrValLysAla----- 61
 Db 9999 CTCACCCGCTGCTGCTTCGTCGCGGCGGCGGCGATGACCGTCGCGCCCGCTGATG 9940
 QY 62 -----LeuAlaSerLysLeuProSerLeuValTyrPheAspPhe 74
 Db 9939 GCGGCTTCTTCATCATGCACTTCGACAGAGTGGCGCGCGCCCTTGGGCTATTTC 9880
 QY 75 AspSerAspGluIleLysProGluAlaAlaIleLeuAspGluGlnAlaGlnPheLeu 94
 Db 9879 GCGGAGGATCGCTTTCACGCGGAGCGGACGACGATCGCATTCGCTGCGCGCATTTGGC 9820
 QY 95 ThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArg 114
 Db 9819 ATTCTGCATTCACCTGCGCAAGATGACCGCGCTGTAGCCCGCG----- 9769
 QY 115 GluTyrAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGly 134
 Db 9768 -----CTCGCGCAAGCGGCGACTC-----ATGCTCGGA 9739
 QY 135 LysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgProIleAla 154
 Db 9738 ATGATTCGCGGAGCGACAGGCTATCATCTGCTTGCCTCGCAGACAGGCGATGATGGCG 9679
 QY 155 Phe 155

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:16:20 ; Search time 1609 Seconds
(without alignments)
311.052 Million cell updates/sec

Title: US-09-674-779b-2
Sequence: 1 MMHIIQIAAANAALSVLFEM.....IAFGNNEAMNSQNRRLASTY 172

Scoring table:
BLOSOM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV-xlp
-O=/cgn2.1/USPTO.spool/US09674779/runat_06072003.121613.9801/app_query.fasta.1.327
-DB=genhmbl -OPMT=fastpct -SUFFIX=2n.rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blcsum62 -TRANS=human40.cdi -LIST=100
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTMT=ptc -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779 -ECGN 1.1.1319.ernat_06072003.121613.9801 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESOURCY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONCLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
Genhmbl:.*
1: gb_ba:.*
2: gb_htg:.*
3: gb_in:.*
4: gb_cm:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_sts:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vl:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_cm:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_sts:.*
28: em_un:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	861	100.0	519	6 AX011034	AX011034 Sequence
2	861	100.0	92407	6 AX067461	AX067461 Sequence
3	855	99.3	519	6 AX011035	AX011035 Sequence
4	850	98.7	519	6 AX011038	AX011038 Sequence
5	844	98.0	519	6 AX011040	AX011040 Sequence
6	316	36.7	7577	1 PEPAL1	X74218 Pseudomonas
7	307	35.7	1048	1 PAOPRTGN	X50191 P. aeruginos
8	306	35.5	13263	1 AE004530	AE004530 Pseudomonas
9	289	33.6	20941	1 AE008730	AE008730 Salmonella
10	289	33.6	294050	1 AL627268	AL627268 Salmonella
11	287.5	33.4	6967	1 ECH297885	AE005987 Caulobact
12	286.5	33.3	10029	1 AE005987	AE005987 Caulobact
13	285.5	33.3	212050	1 AL646050	AL646050 Ralstonia
14	285.5	33.2	7713	1 ECPAL	X05123 E. coli pal
15	285.5	33.2	1531	1 ECXECG	X65796 E. coli exc
16	285.5	33.2	10653	1 AE000177	AE000177 Escherich
17	285.5	33.2	11670	1 AE005252	AE005252 Escherich
18	285.5	33.2	16419	1 AP002553	D90713 Escherichia
19	285.5	33.2	297816	1 AP002553	AP002553 Escherich
20	283	32.9	11138	1 AE013906	AE013906 Yersinia
21	283	32.9	210050	1 AE014146	AE014146 Yersinia
22	276.5	32.1	12393	1 AE012414	AE012414 Xanthomon
23	272.5	31.6	11713	1 AE011958	AE011958 Xanthomon
24	269.5	31.3	349116	1 AE003003	AE003003 Mesorhizo
25	265.5	30.8	10833	1 AE004009	AE004009 Xylella f
26	264	30.7	462	6 AR089419	AR089419 Sequence
27	264	30.7	737	1 HPA15KLP	M18878 H. influenza
28	264	30.7	737	6 I06718	I06718 Sequence 2
29	264	30.7	737	6 I06718	I06718 Sequence 3
30	264	30.7	737	6 I06718	I06718 Sequence 4
31	264	30.7	867	6 HEAOMPP6	M19391 H. influenza
32	264	30.7	867	6 HEAOMPP6	M19391 H. influenza
33	262.5	30.7	13620	1 U32722	U32722 Sequence 1
34	262.5	30.5	340900	1 SME591791	U32722 Haemophilus
35	262	30.4	10906	1 AE009302	AE009302 Agrobacte
36	259.5	30.1	11049	1 AE008312	AE008312 Agrobacte
37	259.5	30.1	10643	1 AE012837	AE012837 Chlorella
38	258	30.0	11914	1 AE004259	AE004259 Chlorella
39	258	30.0	393	1 PMU21673	AJ221673 Pasteurel
40	258	30.0	453	1 AE006136	AE006136 Pasteurel
41	257.5	29.9	11061	1 AE006136	AE006136 Pasteurel
42	254	29.9	237523	1 REXX04	AJ235273 Pasteurel
43	254	29.5	845	1 BRU0016A	L27996 Brucella ab
44	254	29.5	5937	1 AE009476	AE009476 Brucella
45	253.5	29.4	10099	1 AE009476	AE009476 Brucella
46	245	28.6	1213	1 AE009476	AE009476 Brucella
47	245	28.6	669	1 HDU42466	X89009 A. pleuropne
48	243.5	28.3	1540	1 LPLIPO	U42466 Haemophilus
49	241	28.0	10359	1 AE008667	X60543 L. pneumophi
50	215.5	25.0	540	6 AX055138	AE008667 Rickettsi
					L47642 Bartonella
					AX055138 Sequence

BASE COUNT	139 a	108 c	122 g	150 t	/organism="Moraxella catarrhalis" /db_xref="taxon:480"
ORIGIN					
Alignment Scores:	1,96e-82	Length:	519		
Pred. No.:	861.00	Matches:	172		
Score:	100.00%	Conservative:	0		
Percent Similarity:	100.00%	Mismatches:	0		
Best local Similarity:	100.00%	Indels:	0		
Query Match:	6	Gaps:	0		
DB:					
US-09-674-779B-2 (1-172) x AX011034 (1-519)					
Oy	1 MetMetLeuHisIleGlnIleAlaAlaAlaAlaAlaLeuSerValLeuThrPheMet 20				
Db	1 ATGATGTACATATTCAATTGCGCGCGCGCGCGCTTATCGGTACTTAACCTTTTATG 60				
Oy	21 ThrGlycysAlaAsnSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40				
Db	61 ACAGCGCTGTGCCAATTAATCAACAGACAGCATATGTTGCTCTCAATGCACCCACAGCT 120				
Oy	41 TyrThrGlyValIleIleTyrThrGlyValAlaProLeuValAspAsnSpGluThrValLys 60				
Db	121 TACACTGGGGTTACTATCTATACGTGCTTGCACTTTGGTAAATATATGATGAGCCGTTAAG 180				
Oy	61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80				
Db	181 GCTCTGGCAGACAGCACTCCACAGTTGGTTATTTTGACTTGATTCGTGATGACATTAATA 240				
Oy	81 ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100				
Db	241 CCGCAGAGCTGCTGCCCATCTTACAGCAACACACACAACTTTTAAACCACCAACCAACAGCT 300				
Oy	101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120				
Db	301 CGGTGTTGGTGTGACAGGTCATACCCGATGAGCGCTGGTAGTCCGTGACTAATAATATCTACTG 360				
Oy	121 GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140				
Db	361 GGGGAACGCCGTCGCGTGGCGCGGTACGCCAATCTTTTGCTGGTGAAGGCATTAAATCAAGCC 420				
Oy	141 SerValGlnIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGln 160				
Db	421 AGCGTTGAGTTTATCAAGTTTGGTGAAGAACCCTCATGCACTTTGGCACAATATGAGAA 480				
Oy	161 AlaTrpSerGlnAsnArgAlaGluGluSerTyr 172				
Db	481 GCATGCTCACAAAATCGTCTGCTGAACGTCTTAT 516				

RESULT 2	AX067461/c				
LOCUS	AX067461	92407 bp	DNA	linear	PAT 24-JAN-2001
DEFINITION	Sequence 36 from Patent WO0078968.				
ACCESSION	AX067461				
VERSION	AX067461.1	GI:12545081			
KEYWORDS	.				
SOURCE	Moraxella catarrhalis.				
ORGANISM	Moraxella catarrhalis				
	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;				
REFERENCE	Moraxella.				
AUTHORS	1 (bases 1 to 92407)				
TITLE	Lagace, R.E., Peterson, C. and Berg, K.L.				
JOURNAL	Nucleotide sequences of moraxella catarrhalis genome				
	Patent: WO/0078968-A 36 28-DEC-2000;				
	Incyte Genomics, Inc. (US)				
FEATURES	Location/Qualifiers				
source	1..92407				
	/organism="Moraxella catarrhalis"				
	/db_xref="taxon:480"				
BASE COUNT	26788 a 17581 c 20150 g 27888 t				
ORIGIN					

Alignment Scores:

Pred. No.: 1.51e-79
Score: 861.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%

Length: 92407
Matches: 172
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-674-779b-2 (1-172) x AX067461 (1-92407)

QY 1 MettelleuhsitleginllleallalaalaleuSerValleuthrPhem 20
DB 26057 ATGATGTTACATATTCATAATTCGCCGCCGCCGCTTATTCGCTACTTATG 25998
QY 21 ThrlGlySaIaanslySerThrSerGlnValMetValAlaProAsnAlaProThrlGly 40
DB 25997 ACAGGCTGCCCATTAATACACACAGTATGTTGCTCTTAATGACCCACAGCT 25938
QY 41 TyTThrlGlyValleThrlGlyValAlaProleuValAspAsnAspGluThrVallys 60
DB 25937 TACACTGGGTTATCTACTGCTGTTGACCTTGTGATATGATGACAGCTTAA 25878
QY 61 AlaleuAlaserlyseupProSerleuValTyrrheAspPheAspSerAspGluIlelys 80
DB 25877 GCTCTGCAAGCAGCTACCCAGTTGGTTATTTGACTTGTGATTCGATGAGATTAAA 25818
QY 81 ProGlnAlaAlaAlaIleleuAspGluGlnAlaGlnPheleuthrThrAsnGlnThra 100
DB 25817 CCCCAAGCTGCTGCCATCTTACAGCAGACACACATTTTACCACCAATCAACAGCT 25758
QY 101 ArgValleuValAlaGlyHisThrAspGluArgGlySerArgGluTyrrasmetSerleu 120
DB 25757 CGGTTTGGTTGCGAGGTATACCGATAGCGGTGCTGCTGATATATATGTCACCTG 25698
QY 121 GlyGluArgArgAlaValAlaValArgAsnTyrrheleuGlyLysGlyIleAsnGlnAla 140
DB 25697 GGGGAACGCGCTGCGGTGGCTAGCACAATTTGCTGTTAAAGCATTAATCAAGCC 25638
QY 141 SerValGluIleIleSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGlu 160
DB 25637 ACCGTTGAGATTATAGTTTGGTGAAGAACGCCCTATCCGATTTGGCACAAATGAAGAA 25578
QY 161 AlatrPserGlnAsnArgArgAlaGluIleuSerTyrr 172
DB 25577 GCATGTCACAAATCGTCTGCTGACTGCTTAT 25542

RESULT 3
AX011036
LOCUS AX011036 519 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 3 from Patent WO957277.
ACCESSION AX011036
VERSION AX011036.1 GI:997680
KEYWORDS
SOURCE
ORGANISM
Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
1 (bases 1 to 519)
Ruelle,J.L.
AUTHORS
TITLE
Bash019 proteins and genes from moraxella catarrhalis, antigens,
antibodies, and uses
JOURNAL
RUELE JEAN LOUIS (BE); SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES
source
1. 519
location/Qualifiers
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 139 a 108 c 121 g 151 t
ORIGIN
Alignment Scores: 8.5e-82 Length: 519

Score: 855.00
Percent Similarity: 99.42%
Best Local Similarity: 99.42%
Query Match: 99.30%

US-09-674-779b-2 (1-172) x AX011036 (1-519)

QY 1 MettelleuhsitleginllleallalaalaleuSerValleuthrPhem 20
DB 1 ATGATGTTACATATTCATAATTCGCCGCCGCCGCTTATTCGCTACTTATG 60
QY 21 ThrlGlySaIaanslySerThrSerGlnValMetValAlaProAsnAlaProThrlGly 40
DB 61 ACAGGCTGCCCATTAATACACAGTATGTTGCTCTTAATGACCCACAGCT 120
QY 41 TyTThrlGlyValleThrlGlyValAlaProleuValAspAsnAspGluThrVallys 60
DB 121 TACACTGGGTTATCTACTGCTGTTGACCTTGTGATATGATGACAGCTTAA 180
QY 61 AlaleuAlaserlyseupProSerleuValTyrrheAspPheAspSerAspGluIlelys 80
DB 181 GCTCTGCAAGCAGCTACCCAGTTGGTTATTTGACTTGTGATTCGATGAGATTAAA 240
QY 81 ProGlnAlaAlaAlaIleleuAspGluGlnAlaGlnPheleuthrThrAsnGlnThra 100
DB 241 CCCCAAGCTGCTGCCATCTTACAGCAGACACACATTTTACCACCAATCAACAGCT 300
QY 101 ArgValleuValAlaGlyHisThrAspGluArgGlySerArgGluTyrrasmetSerleu 120
DB 301 CGGTTTGGTTGCGAGGTATACCGATAGCGGTGCTGCTGATATATATGTCACCTG 360
QY 121 GlyGluArgArgAlaValAlaValArgAsnTyrrheleuGlyLysGlyIleAsnGlnAla 140
DB 361 GGGGAACGCGCTGCGGTGGCTAGCACAATTTGCTGTTAAAGCATTAATCAAGCC 420
QY 141 SerValGluIleIleSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGlu 160
DB 421 ACCGTTGAGATTATAGTTTGGTGAAGAACGCCCTATCCGATTTGGCACAAATGAAGAA 480
QY 161 AlatrPserGlnAsnArgArgAlaGluIleuSerTyrr 172
DB 481 GCATGTCACAAATCGTCTGCTGACTGCTTAT 516

RESULT 4
AX011038
LOCUS AX011038 519 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 5 from Patent WO957277.
ACCESSION AX011038
VERSION AX011038.1 GI:997681
KEYWORDS
SOURCE
ORGANISM
Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
1 (bases 1 to 519)
Ruelle,J.L.
AUTHORS
TITLE
Bash019 proteins and genes from moraxella catarrhalis, antigens,
antibodies, and uses
JOURNAL
RUELE JEAN LOUIS (BE); SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES
source
1. 519
location/Qualifiers
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 140 a 108 c 120 g 151 t
ORIGIN

Alignment Scores: 2.88e-81
Pred. No.: 850.00
Score: 98.84%
Percent Similarity: 98.84%
Best Local Similarity: 98.26%

Query Match: 98.72% Index: 0
 DB: Gaps: 0
 US-09-674-779b-2 (1-172) x AX011038 (1-519)

OY 1 MetMetLeuHisIleGlnIleAlaAlaAlaAlaLeuSerValLeuThrPheMet 20
 1 ATGATGTTACATATTCATTAATTCGCGCGCGCGCGCGCTTATCGGTACTACCTTTATG 60
 DB 21 ThrGlyCysAlaSerLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
 61 ACAGCGCTGGCCAAATTAATCAACAGCAAGTCAAGTTCAGTTCCTATGACCCACAGGT 120
 OY 41 TyrThrGlyValIleIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60
 121 TACACTGGGGTATCTACTAGTGGTTCACCTTGTGATGATTAATGATGACATATCAAG 180
 DB 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80
 181 ACTTTGGCAAGCAGCAGCTACCCAGTTGGTATTGCTTGTGATTTGATGATGATGAT 240
 OY 81 ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
 241 CCGCAAGCTGCTGCCATCTTACAGCAACAGCAATTTTACACCAATCAAAACAGCT 300
 DB 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
 301 CGTGTGTTGGTTCAGAGTCACTACCGATGAGCGTGGTATGCTGATTAATATATGCTG 360
 OY 121 GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
 361 GGGGAACGGCGGTGCGGTGGCGGTACGCACTATTGCTTGTGATTAATATATGCTG 420
 DB 141 SerValGluIleIleSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGlu 160
 421 AGCGTTGAGATTAATCAAGTTTGGTGAAGAACGCCCTATCGCATTTGGCACAATGAAGAA 480
 OY 161 AlaTrpSerGlnAsnArgArgAlaGluLeuSerTyr 172
 481 GCATGTCACAAATATGCTGCTGCACTGCTTAT 516
 DB

RESULT 5
 AX011040 519 bp DNA linear PAT 06-SEP-2000
 LOCUS Sequence 7 from Patent WO9957277.
 DEFINITION AX011040
 ACCESSION AX011040
 VERSION AX011040.1 GI:9997682
 KEYWORDS
 SOURCE Moraxella catarrhalis.
 ORGANISM Moraxella catarrhalis
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 Moraxella.
 REFERENCE 1 (bases 1 to 519)
 AUTHORS Ruehle,J.L.
 TITLE Babs019 proteins and genes from moraxella catarrhalis, antigens,
 antibodies, and uses
 JOURNAL Patent: WO 9957277-A 7 11-NOV-1999;
 RUEHLE JEAN LOUIS (BE); SMITHKLINE BEECHAM BIOLOG (BE)
 FEATURES
 source 1..519
 /organism="Moraxella catarrhalis"
 /db_xref="taxon:480"
 BASE COUNT 137 a 111 c 122 g 149 t
 ORIGIN

Alignment Scores: 1,25e-80 Length: 519
 Pred. No.: 844.00 Matches: 169
 Score: 98.26% Conservative: 0
 Percent Similarity: 98.26% Mismatches: 3
 Best Local Similarity: 98.03% Indels: 0
 Query Match: 6 Gaps: 0
 DB:

US-09-674-779b-2 (1-172) x AX011040 (1-519)

OY 1 MetMetLeuHisIleGlnIleAlaAlaAlaAlaLeuSerValLeuThrPheMet 20
 1 ATGATGTTACATATTCATTAATTCGCGCGCGCGCGCTTATCGGTACTACCTTTATG 60
 DB 21 ThrGlyCysAlaSerLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
 61 ACAGCGCTGGCCAAATTAATCAACAGCAAGTCAAGTTCAGTTCCTATGACCCACAGGT 120
 OY 41 TyrThrGlyValIleIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60
 121 TACCGTGGCTTAATCACTAGCTGGTGGCACTTTGGTATGATTAATATGACGCCGCAAG 180
 DB 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80
 181 GCTTTGGCAACAGCAGCTACCCAGTTGGTATTGATTTGATGATTTGATGATGAT 240
 OY 81 ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
 241 CCGCAAGCTGCTGCCATCTTACAGCAACAGCAATTTTACACCAATCAAAACAGCT 300
 DB 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
 301 CGTGTGTTGGTTCAGAGTCACTACCGATGAGCGTGGTATGCTGATTAATATATGCTG 360
 OY 121 GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
 361 GGGGAACGGCGGTGCGGTGGCGGTACGCACTATTGCTTGTGATTAATATATGCTG 420
 DB 141 SerValGluIleIleSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGlu 160
 421 AGCGTTGAGATTAATCAAGTTTGGTGAAGAACGCCCTATCGCATTTGGCACAATGAAGAA 480
 OY 161 AlaTrpSerGlnAsnArgArgAlaGluLeuSerTyr 172
 481 GCATGTCACAAATATGCTGCTGCACTGCTTAT 516
 DB

RESULT 6
 PPPAL 7577 bp DNA linear BCT 16-JUN-1999
 LOCUS Pseudomonas putida ruvb (partial), tolQ, tolR, tolA, tolB and oprL
 DEFINITION genes.
 ACCESSION X74218
 VERSION X74218.2 GI:5514773
 KEYWORDS oprL gene; tolA gene; tolB gene; tolR gene.
 SOURCE Pseudomonas putida
 ORGANISM Pseudomonas putida
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 REFERENCE 1
 AUTHORS Rodriguez-Herva,J.J., Ramos-Gonzalez,M.I. and Ramos,J.L.
 TITLE The Pseudomonas putida peptidoglycan-associated outer membrane
 lipoprotein is involved in maintenance of the integrity of the cell
 cell envelope
 JOURNAL J. Bacteriol. 178 (6), 1699-1706 (1996)
 MEDLINE 96198174
 PUBMED 8626299
 REFERENCE 2
 AUTHORS Rodriguez-Herva,J.J. and Ramos,J.L.
 TITLE Characterization of an oprL null mutant of Pseudomonas putida
 JOURNAL J. Bacteriol. 178 (19), 5836-5840 (1996)
 MEDLINE 96422022
 PUBMED 8824639
 REFERENCE 3 (bases 1 to 498)
 AUTHORS Ramos-Gonzalez,I.
 TITLE Direct Submission
 JOURNAL Submitted (21-JUL-1993) I. Ramos-Gonzalez, C.S.I.C. Estacion Exper.
 del zaidin, 18008 Granada, SPAIN
 REFERENCE 4
 AUTHORS Ramos-Gonzalez,I.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUN-1995) I. Ramos-Gonzalez, C.S.I.C. Estacion Exper.

AUTHORS McCelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Portolick, S., All, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.
TITLE Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21534948
PUBMED 11677609
REFERENCE 2 (bases 1 to 20941)
AUTHORS The *Salmonella typhimurium* Genome Sequencing Project.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of Pathology, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/genemark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one mis subclone.

FEATURES

SOURCE Location/Qualifiers
 1..20941
 /organism="Salmonella typhimurium LT2"
 /strain="LT2; SSSC 1412; ATCC 700720"
 /db_xref="ATCC:700720"
 /db_xref="taxon:99287"
 /note="LT2"
 90..568
 /gene="sdhC"
 /note="STM0732"
 90..98
 /gene="sdhC"
 /note="putative -35-signal for sdhC;
 RegulonDB:STMLTH004697"
 115..123
 /gene="sdhC"
 /note="putative -10-signal for sdhC;
 RegulonDB:STMLTH004697"
 166..171
 /gene="sdhC"
 /note="putative RBS for sdhC; RegulonDB:STMS1H001042"
 179..568
 /gene="sdhC"
 /EC_number="1.3.99.1"
 /EC_number="1.3.5.1"
 /note="similar to E. coli succinate dehydrogenase, cytochrome b556 (AAC73815.1); Blastp hit to AAC73815.1 (129 aa), 92% identity in aa 1 - 129"
 /codon_start=1
 /transl_table=11
 /product="succinate dehydrogenase, cytochrome b556"
 /protein_id="AAL19676.1"
 /db_xref="GI:16419242"
 /translation="MIRNKKQRPVNDLQITIRPTTAIASILHRVSGVITFAVGIL

CDS

CDS

gene LMLTSLSPGCFQOADMIDGFLVFMGILFALAHYIVGIRHMLDFGLEET
RBS 549..909
 /gene="sdhD"
 /note="STM0733"
 549..554
 /gene="sdhD"
 /note="putative RBS for sdhD; RegulonDB:STMS1H001043"
 562..909
 /gene="sdhD"
 /EC_number="1.3.99.1"
 /EC_number="1.3.5.1"
 /note="similar to E. coli succinate dehydrogenase, hydrophobic subunit (AAC73816.1); Blastp hit to AAC73816.1 (115 aa), 94% identity in aa 1 - 115"
 /codon_start=1
 /transl_table=11
 /product="succinate dehydrogenase, hydrophobic subunit"
 /protein_id="AAL19677.1"
 /db_xref="GI:16419243"
 /translation="MVSASALGRNGVHDFILVRAIVLITIIYMGFFATSELT
 FEAWTGFSSAFETKVFLLALFSLIHANIGMQLVLDYVKPLANLILQIVVALV
 VYVIGFVYVWGV"
 892..2675
 /gene="sdhA"
 /note="STM0734"
 892..897
 /gene="sdhA"
 /note="putative RBS for sdhA; RegulonDB:STMS1H001044"
 909..2675
 /gene="sdhA"
 /EC_number="1.3.99.1"
 /EC_number="1.3.5.1"
 /note="similar to E. coli succinate dehydrogenase, flavoprotein subunit (AAC73817.1); Blastp hit to AAC73817.1 (588 aa), 97% identity in aa 1 - 588"
 /codon_start=1
 /transl_table=11
 /product="succinate dehydrogenase, flavoprotein subunit"
 /protein_id="AAL19678.1"
 /db_xref="GI:16419244"
 /translation="MKLPVREFDAVVGAGGMAAIOISQGTALSKVPYPTS
 HTVSAAGITVALGNTHEDNMEMHVDYKGSYDYGDDAIEHCKTGPELLEHM
 GLEPFSGLDGRITQRPFGOSKNRGEQAKRTAAADRTGHALHTTYQNALNHTTI
 FSPWYLLDIYKODGAVGCTALCIETEVEYFKARATVATGAGRIYQSTNAHIN
 TGDVGMAIRACVPODMEMQWHPPTGAGAGVLTTECGREGGVLINKRBERMEY
 APNAKDLADRVVARSIMIEIREGCGDGPMPKPKLKHIGREVLNRPGLTELS
 RTEAHDPVKEPIPVIPCHYMGGIPKVTGQALTVNEQEDVYIGIPAVGTACV
 SVGARNPLGSLDLVYGRAGAGLHIGESTIAEGCVLDAESYDESLERLNMWNN
 RNEGDEVLIRKALQECMOHNPVPRFGAMAKGLEQLVIERLKNARLDDTSSEFNT
 ORECELDNLMETATRVASANFTESRGAHSRDEPERDANMLCHTLIYQPTESM
 TRSVNMEPKLRFAPPKIRTY"
 2678..3408
 /gene="sdhB"
 /note="STM0735"
 2678..2683
 /gene="sdhB"
 /note="putative RBS for sdhB; RegulonDB:STMS1H001045"
 2689..3408
 /gene="sdhB"
 /EC_number="1.3.99.1"
 /EC_number="1.3.5.1"
 /note="similar to E. coli succinate dehydrogenase, iron sulfur protein (AAC73818.1); Blastp hit to AAC73818.1 (238 aa), 96% identity in aa 1 - 238"
 /codon_start=1
 /transl_table=11
 /product="succinate dehydrogenase, Fe-S protein"
 /protein_id="AAL19679.1"
 /db_xref="GI:16419245"
 /translation="MMKLEFSIYVNDPNANRMDYTLLEGEGMDMLDALIQLK
 EKDPSLSRSCREGCGSDGINKNGNAGIATTPISALDPEKRTIVIRPLGGLVIR
 DLVYDMGCFYAOYERIKIPYLLNNGONPAREHNEPQREKIDGLYEITLACGCTSC

/aa-2-oxoglutarate dehydrogenase
 /note="similar to E. coli 2-oxoglutarate dehydrogenase
 (decarboxylase component) (AAC73820.1); Blastp hit to
 AAC73820.1 (933 aa), 94% identity in aa 1 - 933"
 /codon_start=1
 /transl_table=11
 /product="2-oxoglutarate dehydrogenase decarboxylase
 component"
 /protein_id="XAL19680.1"
 /db_xref="GI:16419246"
 /translation="MSKSLAKMLADSSYLSGNSQSHITEDLYEDPTDPSDANMRLR
 FOOLPGTGVKFDLDASTRYFRFQRLALASRHSSTISDPTNKKVQYLLQNLNRRP
 GHQNALDPTFVLMKEQVADLADSPFLDELQVETFEFNFGVSAKSEKEMKIGELDLA
 KQYQGFPGAEVYHITSTEERKNLQORIEGSAFSADEKRFNLELQALGLELYEG
 ALKPFKAEKSEGGDALIPMLKEVRIAGNSGTRFEVLLDMARGNLANLIVLQKPK
 DLDPFKAEKHEHGTGDVYVHNGSSDITVEGGLVHALNPNSEHLEPVVWSSG
 RLNDRLDEPSSKYLPTTHGCAAVNGGCVQVETELNMSKAGVEGNGVRLVINOQV
 GETTSPIDANSTYQYDITGDMQAPLFRHNDPDAVAPVIRALDPLDPAETEMVNI
 DLYACRRHGHNEADVEPSATPQPLWQTKIKRPIYKTIADKLEADKATPLDPAETEMVNI
 YDALDAECVAKEMKRPNNHMSFSPYLHNHEDVAPVKEVKEKLOEALAKISTVPE
 ALTEMRSYAKITGROPLAAAGKELFDWGGKMNVALVTLDEGI PVKLSGSDGRST
 IHNAVYHNQTSYTPPLTHHSQSGCFKWDVSLSEEVYLAIEGYATAEERLTITL
 EAPQDPLFNGAOVYDIEISSGEQKMGKGLVMLPHGIEGEPHSSALERTLQILANG
 CAGQAKVQCVSTPQAVQYHMLRQALGMBRLPVVMSPLSHRLPASTVDELINQ
 PGPALGEIDELDPAKAVKRVWVSSKRYITDLEQRNRNDQDAIVAVTEQLYEPFKAV

Alignment Scores:	
Pred. No.:	1,03e-19
Score:	289.00
Percent Similarity:	54.95%
Best Local Similarity:	36.81%
Query Match:	33.57%
DB:	1
Length:	20941
Matches:	67
Conservative:	33
Mismatches:	64
Indels:	18
Gaps:	3

OY	I	MektelkuehisrlleglnlalaiaialaialaleuSerValleuthrPheMet	20
Db	19478	ATGCATTGACAAAGAAGTCTGAAGGCCCTCATGATTCGCCCTGTATGGCAATCGCG	19537
OY	21	ThrclyCysAlaanslySerrthrSergInValMetValAAlaProhsnlaAProThrgly	40
Db	19538	GCATGTTCTCCAACAGACGGCCGCAAT-----GACGGT	19573
OY	41	TyrThrclValIletryrThrIsgValAlaProLeuValAsp-----	54
Db	19574	AGCCAAAGCGGTATGCTGAACGGCGCCGACACTGGTATGAGCCCTTAAGGCCAAAGCCAMC	19633
OY	55	-----AsnspglurThrValysAlaAlaLeuAlaSerlysLeuPro-----SerLeuVal	70
Db	19634	ATGTCACTCATAGAGCAAAGCGCGCTCTGCACATCCAGACGCTCGAGAGAAACAACATCGTT	19693
OY	71	TyrrPheaspPheasPseraspPguIleIysProGlnAlaIalaleuAspGluGln	90
Db	19694	TACTTCGATTCGCAAGACAGTACAGATATCCGTTTCACCTTCGCGGCAATGCTGGATGGCGAC	19755

QY	91	AlaGlnPheLeuThrAsnGlnIrrAlaArgValLeuValAlaGlyHisThrAspGlu	110
Db	19754	GCTAACTCTCCGGTGGCAACCCGGTTTCAAAAGTCAACCGTGAAGAGTCACGGCGACCA	19813
QY	111	ArgGlySerArgGlyIrrAsnMetSerLeuGlyIuArgAlaValAlaValArgAsn	130
Db	19814	CGGGGTACTCCGGAAGTAAACAATCTCCCTGGGTGACCTGTGCTAAACCCCGTTAAATG	19873
QY	131	TyrLeuGlnGlyArgGlyIrrLeuAsnGlnAlaSerValGluIleIleSerPheGlyGlnIu	150
Db	19874	TACTGTGAGGTGAAGGCCGTTTCCGCTGACACAGATTCATCGCTGTTTACAGCTAAAGAA	19933
QY	151	ArgPheIrrLeuAlaPheGlyIrrAsnGlnIuAlaIrrPheSerGlnAsnArgArgIrrGluLeu	170
Db	19934	AAACCTGCGCTACTGTGGCGACGACGAAAGCGGCTTACGCTAAGAACCGTGCACGCTGACTAG	19993
QY	171	SerTyr	172
Db	19994	GTTTAC	19999

RESULT 10			
LOCUS	AL627268		
DEFINITION	Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome, segment 4/20.		
ACCESSION	AL627268	AL513382	
VERSION	AL627268.1	GI:16501953	
	294050 bp	DNA	linear
			BCI 06-JUN-2007

NEIMORUS SOURCE ORGANISM	
	Salmonella enterica subsp. enterica serovar Typhi.
	Salmonella enterica subsp. enterica serovar Typhi
Bacteria:	Proteobacteria; gamma subdivision; Enterobacteriaceae

Salmonella.
1 (bases 1 to 294050)
REFERENCE

AUTHORS

Complete genome sequence of a multiple drug resistant *Salmonella*

JOURNAL Nature 413 (6858), 848-852 (2001)

MEDLINE

PUBMED
REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Notes:
Details of *S. typhi* sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).

FEATURES

nos

den

3.

CDS

Mon Jul 7 08:55:25 2003

```

gene
3594. 4364
/gene="STY0762"
/note="Similar to Mycobacterium tuberculosis hypothetical
3r.3 kDa protein cy13d12.17 rv3783 or mtcyl3d12.17
TR:P72049 (EMBL:Z80343) (280 aa) fasta scores: E():
1.4e-26, 42.2% id in 187 aa, and to Yersinia
enterocolitica O-antigen export system permease protein
RfbD rfbD SR:RFBP_YEREN (Q56902) (235 aa) fasta scores:
E(): 9.3e-19, 29.4% id in 187 aa. This CDS contains a stop
codon at 188. The sequence has been checked and is
believed to be correct"
/pseudo
/codon_start=1
/transl_table=11
/product="putative polysaccharide export ABC transporter
permease protein (pseudogene)"
3603. 4134
/gene="STY0762"
/note="Pfam match to entry PF01061 ABC2_membrane_ABC-2
type transporter, score 21.40, E-value 5.2e-07"
/pseudo
4368. 5078
/gene="STY0764"
4368. 5078
/gene="STY0764"
/note="Similar to Mycobacterium tuberculosis hypothetical
30.0 kDa protein cy13d12.15 rv3781 or mtcyl3d12.15
TR:P72047 (EMBL:Z80343) (273 aa) fasta scores: E(): 0,
43.8% id in 233 aa, and to Yersinia enterocolitica
O-antigen export system ATP-binding protein RfB rFBE
SW:RFBP_YEREN (Q56903) (239 aa) fasta scores: E(): 0,
45.3% id in 225 aa"
/codon_start=1
/transl_table=11
/product="putative polysaccharide export ABC transporter
ATP-binding protein"
4368. 5078
/db_xref="GI:16501957"
/db_xref="SP|PEMBL:Q828D5"
/transl_table="MKISCKNYGVILPIPNSSHSRSFKKTPLOAASGGRICSNNGIE
VALKKIDELFTGNGLALIGHNSGGKTLLPYLGAAYKAPTSRGCEIGRTSLIDPM
MGADKGDTGENIKRLGFLGSKNKINITEDEYSELDQFTKIPIRYTSCMWLR
LGSIETAIINPELLIMDEMVAAGSDSDFRKAKMEINSFTISAGIWMATHDELAKSV
CNKEFIENSELYSGKF"
4359. 5051
/gene="STY0764"
/note="Pfam match to entry PF00005 ABC_tran_ ABC
transporter, score 110.90, E-value 2.5e-29"
4560. 4583
/gene="STY0764"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
5172. 6945
/gene="STY0765"
/pseudo
5172. 6945
/gene="STY0765"
/note="Similar to Mycobacterium tuberculosis hypothetical
71.5 kDa protein rv3808c or mtv026.13C TR:O53585
(EMBL:AU02076) (637 aa) fasta scores: E(): 0.00015, 26.6%
id in 467 aa, and to Mycobacterium tuberculosis
hypothetical 33.9 kDa protein cy13d12.16 rv3782 or
mtcyl3d12.16 TR:P72048 (EMBL:Z80343) (304 aa) fasta
scores: E(): 0.16, 25.6% id in 215 aa, and to Klebsiella
pneumoniae RfB rFbe TR:Q48482 (EMBL:L41518) (297 aa)
fasta scores: E(): 0.3, 23.6% id in 233 aa. This CDS
contains at least one stop codon and frameshift. The
sequence has been checked and is believed to be correct"
/pseudo
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein (pseudogene)"
7138. 7971
/gene="STY0767"

```



```

/db_xref="GI:13424916"
/translation="MAMSSNDAPFATGGRGRGRGRSGALSEINWPLVDVMLVLLI
IFMISAPLTTAGVPLELPKTEAAALQNDDEPTTSTKRDQIIVGEGEIEIPENLAPPI
GAAGGCGKPIFVRADGGATTAIVAAQVAGISNAGFNKIMLTETGSPSTGAAPRDA
OPEGADLPRAP"
complement(5662..6360)
/gene="CC3233"
complement(5662..6360)
/feature="identified by match to protein family HM"
/codon_start=1
/transl_table=11
/product="MolA/TolQ/ExbB proton channel family protein"
/protein_id="AAK25195.1"
/db_xref="GI:13424917"
/translation="MDMAAANFSPFALPMQADPMYKSVTIGLILASGMAVILDK
LFRFQALNRANRFEEOYSGRSLEDVAGEGANDPRALPRMQLAKERDASKGA
ONIALSKMTSLAVVAPSTAEALFATATGLIAPATVAVNKFSTADGKYAGRLEGAD
DSTRIOQRRLERY"
complement(6475..6918)
/gene="CC3234"
complement(6475..6918)
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK25196.1"
/db_xref="GI:13424918"
/translation="MIEPTAGVEGEPHOLPVRIYEDDFTGIYHANYLREGR
SDPFRAGVISHTELAKODTGPATVIRBELDKRARIDDLVFTLEPEREGVRLHVRQ
TTRGDEVLEATVAVACISLSGRPRPPDMLAKLSPLWIPDS"
complement(6915..7781)
/gene="CC3235"
complement(6915..7781)
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25197.1"
/db_xref="GI:13424919"
/translation="MKASSMRILRLTAGLAAGVLAAGLAATSAAGARNALIVOLA
AGLATGALIVGFWFRARMLVMAAGLAVITIMNGVSLEGAKKWALGPOLH
TASIVLPFAAFARAGEGDRVAPALALALLTQHDASSLAVALAALAAVLERP
ROALIPWACVAGLACVAGWEPDLPVAVYEGGLDPTFANPVLGVLAVIMISLP
APRLVYARARPERREALALALALMTGWLGNLFGNVPAPVIGYGAAPILAMGLSLGV
LGDSDQDPPAP"
complement(7771..8811)
/gene="CC3236"
complement(7771..8811)
/feature="identified by match to TIGR protein family HM"
TIGR00635"
/codon_start=1
/transl_table=11
/product="Holliday junction DNA helicase RuvB"
/protein_id="AAK25198.1"
/db_xref="GI:13424920"
/translation="MTRVISEGPHGDIAPADRALRQTLAEFGVGOQAKNLRVFE
AAKGGEDLHYLEPGPLGKTKTLAQIVARELVNFRATGSPVANKPGDLAINTLU
KANDVLEIDELHRSNVNELIYVAMEHDVLDIVIGSPASRSIRIDAPPTLVAAT
RAGMLATPLRDRFGIPRILEFYTPAEVLEHVAHARAGVAPITDDGADLAKARNG
RVAGLRLRRVDFATADGADRIDRAAMALARELVESGICDISDRYLAAMLENGG
GPVGEITLAVIAIAERADVEDVIEPYLMQGFQIORTIRGNACGKATLHLGTPPAP
PGOAGALFDEG"
complement(8808..9425)

```

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:

7.38e-20 Length: 10029
286.50 Matches: 65
58.24% Conservative: 34

```

Best Local Similarity: 38.24% Mismatches: 50
Query Match: 33.28% Indels: 21
DB: 1 Gaps: 4
US-09-674-779b-2 (1-172) x AE005987 (1-10029)
QY 7 11eAlaAlaAlaAlaAlaAlaSerValLeuThrPheMetThrGlyCysAlaAsnLys 26
Db 2740 GTGCGCGCTGGCGCGCGCTTGGTGGC-----GGCTGCGTTCGCGT 2699
QY 27 SerThrSerGlnValMetValAlaProAsnAla-----ProThrGly 40
Db 2698 CCGAAGCCCGACGCCGCTACCGCGCGCGCGCTAGCGGACGCGACCGACCGACCGCT 2639
QY 41 TyrThr-----GlyValIleThrThrGlyValAlaProLeuValAspAsnAsp 56
Db 2638 TACACGCGCGCGCGCGCGCGCGCGGTGACCGCGGTATCTGCGC----- 2594
QY 57 GluThrValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSer 76
Db 2593 GGTTCGTGTCAGGACTTGTGTCATATGCGCGACCGGGTCTATTTCACACCGACGAA 2534
QY 77 AspGluIleLysProGlnAlaAlaAlaIleLeuAspLysGlnAlaGlnPheLeuThrThr 96
Db 2533 TACGTCATTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2474
QY 97 AsnGlnThrAlaArgValLeuValAlaGlnIleThrAspGlnLysGlnLysGlnLys 116
Db 2473 TATTGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2414
QY 117 AsnMetSerLeuGlyGlnAlaArgAlaValAlaValAlaArgAsnLysLeuLysGly 136
Db 2413 AACCTGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2354
QY 137 IleAsnGlnAlaSerValGlnIleIleSerPheGlyGlnLysArgProIleAlaPheGly 156
Db 2353 GTTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2294
QY 157 ThrAsnGlnLysAlaIlePheSerGlnAsnArg 166
Db 2293 TCGAGCGAAGAACGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2264

```

RESULT 13
AL646060 212050 bp DNA linear BCT 07-DEC-2001
LOCUS
DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;
segment 4/19.
ACCESSION AL646060 AL646052
VERSION AL646060.1 GI:17427566
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum.
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
REFERENCE 1 (bases 1 to 212050)
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brothier,P., Camus,J.C., Catolico,L.,
Chandler,M., Choise,N., Claudel,Renaud,C., Cunne,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Stigter,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 212050)
AUTHORS Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CBPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,

laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Bouchere@toulouse.inra.fr
<http://sequence.toulouse.inra.fr/R.solanacearum.html>.
Location/Qualifiers
1. 212050

FEATURES

source
/organism="Rastonia solanacearum"
/strain="GMT1000"
/db_xref="taxon:305"
59. .757
/gene="RSC0556"
/note="RS04909"
59. .757
/gene="RSC0556"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD14086.1"
/db_xref="GI:17427567"
/translation="MALTSAPFWFALGSIILNIVLGSNAVVALAARNPFROROR
ALFWGSAALALRIALVLAIVLKLALPYLKTIGAVLLVYIGIKLSEADEAGSHRP
HDLGMLPRLITLADLVSLDNVAVAAAEKRPPTDVALLVGLGSLIPITIGST
LVALAMRPLVITVLAALIGYLAGMLVTPDADMPQAFPYADVAIGCALIVV
IVGMMGRSLRQA"
951. .1472
/gene="pila1"
/note="RSC0557; RS04908"
951. .1472
/gene="pila1"
/function="structural elements: cell exterior; surface
structures"
/note="Product confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TYPE 4 FIMBRIAL PILIN SIGNAL PEPTIDE
PROTEIN"
/protein_id="CAD14087.1"
/db_xref="GI:17427568"
/translation="MAKSHRRRRRGQGFLLIELMIVVAILGILALAIAPAYDYTI
KSRVYEGIALAGAKALYENNAKOSLSIGASLPSKNSALINDSSIGELAVVY
APAVTEAGANTLVITPYGSSSVFNGNLQAGSTPPAMVLWVCAAGKSMPTVYQRTT
ATLPPKATSECR"
1739. .2245
/gene="pila"
/note="RSC0558; RS04907"
1739. .2245
/gene="pila"
/function="structural elements: cell exterior; surface
structures"
/note="Product confidence : hypothetical
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="TYPE 4 FIMBRIAL PILIN SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD14088.1"
/db_xref="GI:17427569"
/translation="MKSRLINKRVQKFTLLIELMIVVAILGILALAIAPAYDYTVR
ARTEGISTAAOKALYSENAKOSLSIGSSVFPPTKNSALINDSSIGELAVVY
TYTTAAGGTIALVPTNAGTALPVSSAPSGPIMMTCYAKNKAASVAPSGTMSLAA

gene
CDS

KYPAEGR"
2351. 4060
/gene="RSC0559"
/note="RS04906"
2351. 4060
/gene="RSC0559"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD14089.1"
/db_xref="GI:17427570"
/translation="MLMPVYALVACNSVPLVAVSHSYPIPTFSEFVAALIGWVALA
GVLGTSWHSKACLPVYVAPLALVIGLVIVLVVNTNPPTFAIVELIGTAVAGL
GACRDVPGVLESIAVAVITGLVIAIECHLPRVSGSLPNTSTIMPDSGRMMGN
LNQPNVATVYLAFLACLEIGSTRRRVAPLALALIGMALTVSRMSMLITVY
GVAGLAMSAREGARBMIRAGVPVIGLAVYOLCNMVAIVANLMLHDLIPISIDEL
OOGVGLRVFLMKHAMHMLAHPWLGSGGDVAMNOYVOTDVLGVHESMAHNLVLDL
LAKVYVGLLAVMLPELGLVHAAMKRRMTPALAFYAVILTVASHSMLEYPLHLYFL
LPFAVAGVYDPRKSRVLSPTNATVITGVATCGAVLGRMMIDYQSEVERLYSLDGP
VELQYRSGLLPLVPTGMLSTANNAGMTAEAPIMAAVEHQAVOFPQSGTQRMAL
ALAFQKTEALTVQRRLHNOYVIDIAGSKLITVCTKKLGLATCARLAKENLVV
GVD"

gene

CDS

complement(4289. .4765)
/gene="moc"
/note="RSC0560; RS04905"
complement(4289. .4765)
/gene="moc"
/function="small molecule metabolism; biosynthesis of
cofactors; carriers; molybdopterin"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MOLYBDENUM COFACTOR BIOSYNTHESIS
PROTEIN C"

gene

CDS

/protein_id="CAD14090.1"
/db_xref="GI:17427571"
/translation="MPLHHPDRAGQAHMVDGDKAVTHRVAVAGLIMLPPTFALV
RDGTAKRGDVGIAIRIAIOGAKRTSLDPLCHPLRIATKAVDFEIDEDADRYRQTVR
AETHQGTVEEMALFAVOGLITLTYDKCAVDKQVIGDVLMEKKGSGSEVNAQ"
4859. .6544
/gene="RSC0561"
/note="RS04904"
4859. .6544
/gene="RSC0561"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD14091.1"
/db_xref="GI:17427572"
/translation="MTLSFRPRARKVTAVALMTWLAAPVSLAOPASPAPAPACA
AAAEVQSNLNSVOIGROSPYIQDADITIAAPYEIPDMGDPSTALSPMEBRIGD
RVQRQIRDPDYPVDPPLSLDNDIGVRIEAKRQHYAGTSSSETEIARVDPGIN
AFALPGYIGVNTVPLVATESSELSAVLGEHIGVLDORHILARKIDSGSEMTALAS
ILAGLAATKSGDAOLAVGQAAVAVSNOIAFSGAREADRVGFTLLTGAGTNPDS
MPFFRRLRVSSIADTGAVPVYARTHPLTGBRIADMEDRRARGLPHPQDPQRPATGFA


```

/function="membrane; Colicin-related functions"
/note="0421: 100 pct identical to TOLB_ECOLI SW: P19934;
alternate gene names cim, excC, lky"
/codon_start=1
/transl_table=11
/product="membrane spanning protein, required for outer
membrane integrity"
/protein_id="AAC73833.1"
/db_xref="GI:1786960"
/translation="MSKATFQNDKLRKRAIISAVLHVLFAALIWSSFDENIASAGG
GGSSIDAVMDSCAVVEOYRKMOSESSESSARKMEQQAEEERKQAAEQER
LKOLEKRLAAEOBKAEAKOAEKOEKOEKAEKAEKAEKAEKAEKAEKAEKAEK
KKAADKARKKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEK
KAAEKAEKRAAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEK
AKKAAEKRAAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEK
ADINNAQOIKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEK
AAKLAKIPKPSQAVVEYFKNAPLDFKP"
3525..4817
/gene="tolB"
/note="b0740"
3525..4817
/gene="tolB"
/function="factor; Colicin-related functions"
/note="0430: 99.8 pct identical to TOLB_ECOLI SW: P19935"
/codon_start=1
/transl_table=11
/product="periplasmic protein involved in the
tonb-independent uptake of group A colicins"
/protein_id="AAC73834.1"
/db_xref="GI:1786961"
/translation="MKQALRVAFEGFLIMASVLAHVRIVDSGVDSGRPIGVFPQW
AGGAPEDIGIVADLRNKGKFNPLDARLPQOGSAGEVQAPAMSLGIDAVVG
QVTPNDGSSYNVAYOLYDTGAGATYLAQNSKVKMKOMLRVGHASDEFEKLGIK
GARTKATVYQTNAGCOFYELRVSQYDNGVNPVYRSGNAPSPDGSKLAVVT
FESGRALAVIQTNAGAVKQVASFPHNKAPAFSPDGSKLALSKTGSINLYMDIA
SQQIRQVTDGRSNMTEPTWFPDSONLAFISDAGRQVTKVINGGAPORTTWESQD
QDADVSDDGKFMVMSNGQOHIADKODLATGQVLSSTFIDETPPLANPNTWITS
SSQMGSVNLVSTDCRFKRLPATDGVKPFAMSPYL"
4668..4694
/gene="tolB"
/note="factor Sigma70; predicted +1 start at 778139"
4789..4816
/gene="tolB"
/note="factor Sigma70; predicted +1 start at 778261"
4852..5373
/gene="pal"
/note="b0741"
4852..5373
CDS
Alignment Scores:
Pred. No.: 1.02e-19 Length: 10693
Score: 285.50 Matches: 67
Percent Similarity: 56.90% Conservative: 32
Best Local Similarity: 38.51% Mismatches: 72
Query Match: 33.16% Gaps: 3
DB: 1
US-09-674-779b-2 (1-172) x AE000177 (1-10693)
QY 1 MetMetLeuHisIleGlnIleAlaAlaAlaLeuSerValIleuThrPheMet 20
DB 4852 ATGCAACTGACCAAGAGGCTGATGATTCCTTCCTGCTTTAGGCAATTGGC 4911
QY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
DB 4912 GCATGCTTCTCCACAAACAGCCAGCAATGACGCGAAGCATGCTGCGCGGC 4971
QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60
DB 4972 --ACTGCTATGATGCGACAGCGCGCAACGCAACATCTTCGGAAGACGCTCGT 5028
QY 61 AlaLeuAlaSerLysLeuPro-----SerLeuValTyrPheAspPheAspSerAspGlu 78

```

```

DB 5029 CTGCAATGCAACAGCTGACGAGCAACACATGCTTACTGATCTGACACATGACAT 5088
QY 79 IleLysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGln 98
DB 5089 ATCCGTTTCGACTTCCTCAATAGCTGATGACATGACATGCAACCTCTCGTAGCAACCG 5148
QY 99 ThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMet 118
DB 5149 TCTTACAAAGTCACTCAGCTAGACGTACGCGGAGCAACGTGTAATCCGAAATACCAATC 5208
QY 119 SerLeuGlyLysLysArgGlnAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyLeuAsn 138
DB 5209 TCCCTGGGTAACTGCTGCGAAGCCGCTTAAGATCTACTGAGGCTAAGGCGCTTCT 5268
QY 139 GlnAlaSerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsn 158
DB 5269 GCAGACCAGATCTTCATCGTTCTTACGCTGAAGAAACCTGCAGTACTGCTCATWGAC 5328
QY 159 GluGlnAlaTrpSerGlnAsnArgAlaGluLeuSerTyr 172
DB 5329 GAAGCGCATCATCTCCAAACACGTCGCGTACTGCTTAC 5370

```

RESULT 17

AE005252

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

misc_feature

gene

CDS

/organism="Escherichia coli O157:H7 EDL933"

/strain="EDL933"

/serotype="O157:H7"

/db_xref="taxon:155864"

/note="enterohemorrhagic"

/note="O-157:H7"

/note="O-157:H7"

/note="O-157:H7"

/note="O-157:H7"

/note="O-157:H7"

/note="O-157:H7"

/note="O-157:H7"

/note="O-157:H7"

/note="O-157:H7"

/note="O-157:H7"

```
gene
/translation="MNCSEIRLKSTEMHLLAKTTTSQCKLEVDHNYLNTLNQVFNK
LRLMTLYDTH"
complement(454..876)
/gene="20897"
complement(454..876)
/gene="20897"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55066.1"
/db_xref="GI:12513663"
/translation="MIPQISEFNLPAVATVASEFTYDTQKESDILYSACCSLKDAY
VSRCLASIHPTLAKNIYETENLSDOKILHDCRSTNEIKTIFFGQOREGLIETSDT
VFSILSSILDSAIKSFQHOIPVNGSVARELLIYDYD"
complement(958..1902)
/gene="20898"
complement(958..1902)
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55067.1"
/db_xref="GI:12513664"
/translation="MSPDNIGSAVNRSPGSPNSDLTKDEKMTSPYOVYIKNTAI
LDPRIVAGIDLTLTOTRETFSESQOLISALOTLHVMLNLTENRREBSYTLAE
YNDTSRTAYQANNEKPCGCGCGLIODEDDNLKIKTFEARIHVAQVQSLAE
EFMFESYSAQONLEPHLLPAPVPLKTEPRTQEDVLLREFKMDNFKLISDE
DICEFVERDIPSPYRFFKFTNLALVDTLQTLANKIL"
complement(2077..2118)
/gene="20899"
complement(2077..2118)
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55068.1"
/db_xref="GI:12513665"
/translation="MSTDRKPVMLFH"
2705..4276
/gene="cyda"
/note="20900"
2705..4276
/function="enzyme; Energy metabolism, carbon: Electron
transport"
/note="Residues 1 to 523 of 523 are 100.00 pct identical
to residues 1 to 523 of 523 from Escherichia coli K-12
Strain MG1655: B0733"
/codon_start=1
/transl_table=11
/product="cytochrome d terminal oxidase, polypeptide
subunit I"
/protein_id="AAG55069.1"
/db_xref="GI:12513666"
/translation="MMLDIVELSRQRLTMYIFLEVPILTGNAFLIATMEVYVLS
GKOYKDMTKRWKGLFIGNFALGVATGTLMEFOGTWMSYSHVSDIFAPALIEL
NMFLESTFVGLFFFGWDRLGKVOHMCVTWLVAGSNLSALMLVANGMNONPLASF
NEIRKMEVSESELVNLNPAQVKEFVHTVASGVYGMFTIGISAWMYMLGRDFAPK
NSEALASFGMAAVLSYVLGDESGYEMGDVOKRLAAIEAEMETOPAPAPYTLFGLP
DOEETKFAIQAIPALGIATRSVDTPVIGIKRLAQOHEERIENGKATSLLEOLRS
GSTDAVNDQNSMKDKGGLLKRYTPNVAADTEAOIQATKDSIPRAVAPLYEER
IMVAGFLLAILALSPSVIRNIGKKMLRALALGIPLPWIAVEAGFVAEYERQ
PWATGEVPTTVANSSILAGDLIFSMVLIGLVTLLFVAFILFMFKFARLGPSSLTKG
RYHFEQSSSTTTPAR"
4292..5431
/gene="cydb"
/note="20901"
CDS
4292..5431
/function="enzyme; Energy metabolism, carbon: Electron
transport"
/note="Residues 1 to 379 of 379 are 100.00 pct identical
to residues 1 to 379 of 379 from Escherichia coli K-12
Strain MG1655: B0734"
/codon_start=1
/transl_table=11
/product="cytochrome d terminal oxidase polypeptide
subunit II"
/protein_id="AAG55070.1"
/db_xref="GI:12513667"
/translation="MIDEVLRFTIMWLLVGVLLIGFVATDGPDMGMLTRELGRNDT
ERRRIINTEAHMNGNOMVLLTAGALFAPAMPYVAAAFSGFYVAMILVLALEFRPV
GEDRSKTEERHRNRMDMGIFESFVPLVIGVAFGNLQGVFPVADLTLYRYTGN
FPQNLNPGGLAGVSVGMITTGATVYLQKRTVGEHLRTIRATAQVAALVTLCFALA
GVWVYIGIDGVVAKSTMDHYAASNDLKEVVRAGAVLVNFTPLMALIPALGVYLP
LTLITLARMDAKAAWAFVSSLLTACILITAGIAMFVPSSTMMNASLTMMDATSSQ
LTLNVTWVAVAVLPILILYTCMYKMFGRITKEDIERNTHSLY"
5559..5852
/gene="ybgE"
5559..5852
/note="20903"
5559..5852
/function="orf; Unknown function"
/note="Residues 1 to 97 of 97 are 100.00 pct identical to
residues 1 to 97 of 97 from Escherichia coli K-12 Strain
MG1655: B0735"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55071.1"
/db_xref="GI:12513668"
/translation="MSKIIATLTYAVVMDKRPDLALSFVALLIAGCMFDPSPRAKTS
ELEIHWGILLMAVCAVGHGVGRFQKVLMOGIFCPPLADIVLIGLIFFF"
6002..6406
/gene="ybgC"
6002..6406
/note="20904"
6002..6406
/function="ybgC"
/note="Residues 1 to 134 of 134 are 100.00 pct identical
to residues 1 to 134 of 134 from Escherichia coli K-12
Strain MG1655: B0736"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55072.1"
/db_xref="GI:12513669"
/translation="NNTILFRPVRYIEDTDAGGVYHASVAFERARTEMLRHHH
FSQALMERVAFYVRKMTVEIYAPARIDMLLETQETTSKGTSLVFQRIYNAENT
ILNEAEVLVCCYDPLMKRALPKSIVAEFRQ"
6403..7095
/gene="colQ"
6403..7095
/note="20905"
6403..7095
/function="colQ"
/note="Residues 1 to 230 of 230 are 100.00 pct identical
to residues 1 to 230 of 230 from Escherichia coli K-12
Strain MG1655: B0737"
/codon_start=1
/transl_table=11
/product="inner membrane protein, membrane-spanning,
maintains integrity of cell envelope; tolerance to group A
colicins"
/protein_id="AAG55073.1"
/db_xref="GI:12513670"
/translation="MTDMNITLDLFLKASLLVKLIMLIIIGFSIASAIIORTIRLNA
AAREAFEDKFWGSGIELSRLOYSGKRNLTGSGQIRYSGFKEVRLHRANSHAPE
AVVEGASRAWRISMNRELETHIIFPLGVGSISPTIGLFTGVGIMHAFIALGAVK
QATLQWAPGAIARLITATIGLFAAIPAVAAVNRILNRQVKNLNLVYDNFMEEFTAILH
```


gene
1168 . 3801
/gene="ybgp"
1168 . 3801
/note="ORF_ID:0178#1
similar to SwissProt Accession Number P54746"
/transl_table=1
/product="Hypothetical protein in hsa 3' region ."
/protein_id="BA35398.1"
/db_xref="GI:4062319"
/translation="MKAIVRYHTIRPHMREMYFTTESRILLVNMMEILCREOD
NEKYVYLDGOTALEDEYPAVKNPKRKYOVAGLIIIGPYTQTDPTIASBIV
RLIMVGMDCARSGEPKIGILPDSFGMSQLEPHIYNGEITRTMRCGSEHGTDK
TEFLMQSDSDSEVTAQVLPGLGKYLPADENGILRLDSDVLEKASVKEILLP
NGHDMPLQONIFEVMDKREIYQKRFVSEFEVEKEIEAQRONLATLKGFEIDLG
YMRVHRTIGSTRMDIKIAHARIEKIVNLEPLATLMTLGPYHGLGKMKKELIK
NHAHDSIGCCSDKREIYAREFELADMDNLTIRYMKRIDNMFQSDADLVLENT
MPREVEINTVYLRASQFNLRDROQPVYITRIAREIDPGLIDRQIVHGNDFP
MERDIOINQIVPSMGRTLYTEANQPEYNVIAASDAEGLIENAFQIALNEGSLQIV
DKDGCYRIRYQLIESSDDGDEYDSPAEEWYITANAAPQCDIIRHMOGRVIR
YDMAVPLNLSERASROETGVVLYVTLSSHNSRIDVDINLDNADDRLEVLVDFP
FNDISVADTQFGSLTRPVNDSAMNMNOOGKEAPVPMNINLYALDOEGNMAVF
SEGLREFEVIGEEKKTPALTLRGVGLIGKEDLLRPGSSGKMPVDSQRLGSLC
RLSLSTYGTPTAGYACQARALTYOCCNKLPMVMKLNKAGEVNPSSYLLMMP
VGCILSLAKKAEDROEYTLKLFNPAASATCATATVAREVISETMDHEHITTENO
GSMISGPFLEQSTKFSYRLA"
4648 . 6216
/gene="cyd1"
/note="cydA"
4648 . 6216
/gene="cyd1"
/note="ORF_ID:0178#2
similar to SwissProt Accession Number P11026"
/codon_start=1
/transl_table=1
/product="Cytochrome d ubiquinol oxidase subunit I (EC
1.10.3.-)."
/protein_id="BA35399.1"
/db_xref="GI:1651327"
/translation="MDIIVELRLQFALTAMVHFLFVPLTGMALFIMETVYVLS
AFPLESTFYGLFEFGMDRLGKVOHMCVTMLVAGSNLSALNLTIVANGMPLIEGLM
KOIKDMTKFMGKLFINFLVATGILTEPQFGNTPVYSSYVGDIGDIALEGMA
FETMRKENVSESELYINPYAQYKPVHTVASGVTGAFTLIGISAMYLKGFDPFARK
SEFIAASRGMAYLSTVIGDSGYEMGVOVTKLALEAMETQPAAPAAFTLGIIDP
QSEETKFAIQLPYALGIIATRSVDTPVIGLKEIMVQHEERIRNOMKAYSLLELRSG
STDQAVRDQFNSMKRDLGIGLLKRYTNVADATQAOIQATKDSIPVAPLYAFPI
MYACGLIATLALISFWSYIRNRIGEEKMLRAALYGLPLPWIAVEAGMFAEYGRDP
MAIGEVLPAAVANSILTAGDLIFSMVLICGLTYTLFVLELTMFRANLGPSSLTGR
VHEOSTTTOPAR"
6232 . 7371
/gene="cyd2"
/note="cydB"
6232 . 7371
/gene="cyd2"
/note="ORF_ID:0178#3
similar to SwissProt Accession Number P11027"
/codon_start=1
/transl_table=1
/product="Cytochrome d ubiquinol oxidase subunit II (EC
1.10.3.-)."
/protein_id="BA35400.1"
/db_xref="GI:1651328"
/translation="MIDEVLRFTIMWLLVGVLLIGFAVTDGFDMGVMLRFLGRNDT
BRKMINIAPHMDGNQVMLITAGGALFAMPVYAAAFSGFVAMTLVYASIFPPY
GEDYRSKILETRMRNMMDGIFIGSPPLVIGVAFGNLLOGVPENDVLRITXYGN
FEOLLNPGGLAGVYVGMITOGATYLOMRTVGEALTLPRATQVAAALTYTVGFALA
GVWVWYGDYGVKSTMDHVAASNPMLKEVVRREGALVFNNTPIIMALPAGVLP
LTLITLARMKAAAFVPSSTLACTITLPGAIAPFPVMSSTMMNASLTMMDATSSQ
LTINMTVAVVAVLPITLITAMCYMKMFRITKEDIERNTHSLY"
7499 . 7792
gene

CDS
/gene="ybgp"
7499 . 7792
/gene="ybgp"
/note="ORF_ID:0178#4
similar to SwissProt Accession Number P37343"
/codon_start=1
/transl_table=1
/product="Hypothetical protein in cydA-toI0 intergenic
region precursor."
/protein_id="BA35401.1"
/db_xref="GI:4062320"
/translation="MSKTIADLYAVMDKRPRLASPMVALLIAGCMFMPSPRAKTS
ELEIWHGLLMMVAVCAAGVINGVFRQKVLMOGIFCPDLADIVLYGLIEFFP"
7942 . 8346
/gene="ybgc"
7942 . 8346
/gene="ybgc"
/note="ybgC"
/note="ORF_ID:0178#5
similar to PIR Accession Number A25980"
/codon_start=1
/transl_table=1
/product="15.5k protein (tolA operon 5' region)."
/protein_id="BA35402.1"
/db_xref="GI:4062321"
/translation="NMTTFRMPVRYEDTDAGVYVYHASYVAREARKDMLRHH
FSOOLMAERVAFVRKMYEYVAPARLDDMLEIOTETISMGTSLVFTORIVALENT
LNEPAVLYVCVPLKMKPRALPKSIVAEFK"
8343 . 9035
/gene="f11"
/note="f10"
8343 . 9035
/note="f11"
/note="f11"
/note="ORF_ID:0178#6
similar to PIR Accession Number B25980"
/codon_start=1
gene

Alignment Scores:

Pred. No.: 1,77e-19 Length: 16419
Score: 285.50 Matches: 67
Percent Similarity: 56.90% Conservative: 32
Best Local Similarity: 38.51% Mismatches: 72
Query Match: 33.16% Indels: 3
Gaps: 2

US-09-674-779b-2 (1-172) x D90713 (1-16419)

QY	1	MeMeLeuHisIleGlnIleAlaIleAlaIleAlaIleuSerValIleuThrPheMet	20
DB	12257	ATGCACACTGACAAAGTCTGAAAGCGCTGATGATGCTGCGCTTATGCAATTCGG	12316
QY	21	ThrGlyCysAlaAspLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly	40
DB	12317	GCATGTTCTTCCACAGACAGACCCAGCAATGACGCAAGCCATGCTGGTGGCGGC	12376
QY	41	TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrVal	60
DB	12377	---ACTGTATGATGAGCGAAGCGCGCAACATGCTTCCGAAAGCAGGCTCGT	12433
QY	61	AlaLeuAlaSerLysLeuPro-----SerLeuValTyrPheAspPheAspSerAspGlu	78
DB	12434	CTGCAAATGCACACAGCTCAGCAGCAACATCGTTACTTCGATCTCGACAAAGTACGAT	12493
QY	79	IleLysProGlnAlaIleAlaIleAlaIleLeuAspGluIleAlaGlnPheLeuThrThrAsnGln	98
DB	12494	ATCCGTTCTGACTTGGCTCAATGATGGATGGATGCACATCTCCGCTAGCAACCG	12553
QY	99	ThrAlaIleValIleuValAlaGlyHisThrAspGluIleGlySerIleGlyIleuThrAsnMet	118
DB	12554	TCTTCAAAAGTACCGCTGAGAGGTACAGCGGACCAAGTGTACTCCGGAATACCAATC	12613
QY	119	SerLeuGlyIleuAlaGlyAlaValAlaValIleAsnTyrLeuLeuGlyIleuGlyIleuAsn	138
DB	12614	TCCCTGGTGAACGTCGTGCAACGCGCTTAAGATGTACTGCAAGGTAAAGCGCTTCT	12673

/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEHVORROCKLEQVDTRVAAIEKGIILVETGNGK
KSTRACTVTRANGHCKRTGVQVYIKQDNGEYMLDPLIGVHMGCTWETONR
QADIDAKEWSESKRMADKRDVLVDDELTYMLAYHLDTEVIASTLORRPAQSV
IVRGCHSQILKMDADTSEIRPVKHAFDNGIQAGPGLDW"
complement(4332..5984)
/gene="ECs0757"
complement(4332..5984)
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
g1120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"

/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPFIMQEPLOKDCETYLISDQHTITELDESEVIRIAE
ALTILARQAFYASPELRSALHQVASTLNDQASNDKYVALQLNMAVSAKVLP
NCOPTGTAIVASKKQOITWGNDAEALSKGISTYEQENSLDPEKALFLYIEVNTQ
TNLPQOITISAVAGDEYHELCVNGGSSANKALYQETSKLQPEKALFLYIEVNTQ
GTACPPYHIAVGGISADQTLIAKASTKYDNPITSGNEGQAPRIETELKXVL
EASQPGVGAQGGKGFVADIRIVRLPHRGGSCP IAAALSCSADRINKAKINHGVL
KLEHNPQVYIPASIRENHOAYOLDRLRDYMDLARLPVGTAVSLSGPTVAV
DIAHAKIKARLDGEMPEYKHHIYVYAPAKTPENMAKGSIGPTGGMDYIDT
OAGSLVIAISKNGRSQVADACHKRGFGNIGSIGGAALAEYKSLCEYPELG
MEAVMMVEVETLPAILVDKGNFSEQFEQHRCAQSPAGH"
complement(6092..7372)
/gene="ECs0758"
complement(6092..7372)
/gene="ECs0758"
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
8a, also similar to C4-dicarboxylate transpor"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKRSILTMILLALVGLIIVLNNTASPEAKLYAQEISIT
TIFRLIKMILAVLVSTLVGIAKKMDAKCRIPSKTFLPICASLITGLITV
NFEPGTGINFVAHGAETGVVAAPETLKVFSHPSTYDAAHNEITLQIVEST
FLGCSLTAIGKSAIVVHALDSLAAHMLKTVMLFALVFAAISALIAERGLAV
VSAIGIFGGEYFMILLVILITGLAIVGPCIIRLTRALSPALALTTSSSEAP
CTLEKEQREVSRIKSFVPLIGISFNLSMAVCSFAVFAQACNHLSTGEIDTM
LILMLTSSKMAVSPASVAVIALNCPINPEAGLILMGVDPPLDMGRSATVNTSN
AMGAAMVSRMEGHGEGRCRALKAPNESNALP"
complement(7533..7850)
/gene="ECs0759"
complement(7533..7850)
/gene="ECs0759"
/note="unknown"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB34182.1"

Alignment Scores:
Pred. No.: 7, 29e-18 Length: 297816
Score: 285.50 Matches: 67
Percent Similarity: 56.90% Conservative: 32
Best Local Similarity: 38.51% Mismatches: 72
Query Match: 33.16% Indels: 3
DB: 1 Gaps: 2

US-09-674-779b-2 (1-172) x AP002553 (1-297816)
QY 1 MetKleUhis1IleClnIleAlaAlaAlaAlaUleUerValIleUthrPheMet 20
Db 24093 ATCCAACTGACCAACAACTGCGTAAGAGGCGTATGATGCTCTGCGCTTATGCAATTCGC 24152
QY 21 ThGlyCysAlaAsnUserThrsGlnValMetValAlaProAsnAlaProThrGly 40
Db 24153 GCATGTTCTTCCACAGAACCCCGCAATGACGCGACGAGGACATGCTGGTGGCCGC 24212
QY 41 TyrThGlyValIleIleTyrThrGlyValAlaProUleValAspAsnAspGluThrVal 60
Db 24213 ---ACTGATGATGATGCGAAGCGCGGCAACGCGCAACATGCTCTCCGAGACAGGCTGT 24269
QY 61 AlaUleAlaSerUleUPro-----SerUleValTyrPheAspPheAspSerAspGlu 78
Db 24270 CTGCAATGCAACAGCTCGACGACGACCAACATCTGTTACTTCTGATCGACAACTACGAT 24329
QY 79 IleUysProGlnAlaAlaAlaIleUleUasPglUglInaIleUthrPheUthrAsnGln 98
Db 24330 ATCCGTTCTGACTGCTGCAAAATGCTGATGATGATGATGATGATGATGATGATGATGAT 24389
QY 99 ThrAlaArgValUleValAlaGlyHisThrsPglUargGlySerArgGluTyrAsnMet 118
Db 24390 TCTTCAAGTACCGCTGAGAGGTCACCGCGGACGACGATGCTGCTCGGAATACACATC 24449
QY 119 SerUleUglUargArgAlaValAlaValArgAsnTyrUleUglUysGlyIleAsn 138
Db 24450 TCCCGGGTGAACGCTGCGGACGCGGATTAAGTGTGACTGAGGATTAAGCGCTTCT 24509
QY 139 GlnAlaSerValGluIleIleSerPheGlyGluUleUargProIleAlaPheGlyThrAsn 158
Db 24510 GCAGACGATCTCATCTGTTTCTTACGTAAGAAAMAACTGCATGATGCTGATGAC 24569
QY 159 GluUleAlaTpsSerGlnAsnArgArgAlaGluUleUerTyr 172
Db 24570 GAAGCGATCTCATCTGTTTCTTACGTAAGAAAMAACTGCATGATGCTGATGAC 24611
RESULT 20
AE013906/c 11138 bp DNA linear BCT 26-JUL-2002
LOCUS
DEFINITION
Yersinia pestis KIM section 306 of 415 of the complete genome.
ACCESSION
AE013906 AE009992
VERSION
AE013906.1 GI:21959961
KEYWORDS
Yersinia pestis KIM.
SOURCE
ORGANISM
Yersinia pestis KIM.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
REFERENCE
AUTHORS
Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F.,
Liss P., Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
Fetherston J.D., Lindler L.E., Brubaker R.R., Plana G.V.,
Straley S.C., McDonough K.A., Nilles M.L., Matson J.S.,
Battner F.R. and Perry R.D.
Genome Sequence of Yersinia pestis KIM
J. Bacteriol. 184 (16), 4601-4611 (2002)
PUBMED
12142430
TITLE
JOURNAL
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
FEATURES
source
1..11138
Location/Qualifiers
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"

```

gene
177..512
/gene="psif"
/note="y3047"
177..512
CDS
/gene="psif"
/function="phenotype; central intermediary metabolism:
phosphorus compounds"
/note="residues 1 to 111 of 111 are 54.46 pct identical to
residues 7 to 112 of 112 from E. coli K12 : B0384;
residues 1 to 111 of 111 are 53.57 pct identical to
residues 1 to 106 of 106 from Genpept : >gb|A019338.1|
(AE008713) induced by phosphate starvation [Salmonella
typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="putative starvation-inducible protein"
/protein_id="AA086597.1"
/db_xref="GI:21959962"
/translation="MRLPMLLSAGLLSANWADPPKTPSPAAQAOQOMTCNO
OASTSLKDERKNMSOCLKAQYAPDGKALTYQOQKMSCNBAQKMLGDERKTF
MSTCLKRAA"
/gene="gpmA"
775..1611
/gene="gpmA"
/note="y3048"
775..1611
CDS
/gene="gpmA"
/function="enzyme; energy metabolism, carbon: glycolysis"
/note="residues 29 to 278 of 278 are 87.99 pct identical
to residues 1 to 250 of 250 from E. coli K12 : B0755;
residues 29 to 278 of 278 are 89.99 pct identical to
residues 1 to 250 of 250 from Genpept : >emb|CAD05220.1|
(AL627268) phosphoglycerate mutase 1 [Salmonella enterica
subsp. enterica serovar typhi]"
/codon_start=1
/transl_table=11
/product="phosphoglyceromutase 1"
/protein_id="AA086598.1"
/db_xref="GI:21959963"
/translation="MRLTERTORTCSIKYDGHIPNLGVWMAVTKLVLRHGESOM
NENRTGWYDVLSERKSEKAAKGLIKDEGPPDPATSYLKRALITMNLIDL
DOAMLPTEKTMKLNENHYGALOGLNKSEAEKYGDQYKORRGPAITPRLPELSDR
PFGHDPYRAKLTDAELPTTESLATIERTVPMNVNIPKASGERVITAAHGSLSLA
LVKYLDELGEDELEINIPTEGVPLVYEEDENKPKTKHYVLTGNADETIAKAAVAVNOK
AK"
complement(1793..2845)
/gene="arog"
/note="y3049"
complement(1793..2845)
CDS
/gene="arog"
/function="enzyme; amino acid biosynthesis: Phenylalanine"
/note="DAH synthetase, phenylalanine repressible;
residues 1 to 349 of 350 are 86.24 pct identical to
residues 1 to 349 of 350 from E. coli K12 : B0754;
residues 1 to 349 of 350 are 87.10 pct identical to
residues 1 to 349 of 350 from Genpept : >emb|CAD05217.1|
(AL627268) phospho-2-dehydro-3-deoxyheptonate aldolase
(DAH synthetase) phenylalanine repressible [Salmonella
enterica subsp. enterica serovar typhi]"
/codon_start=1
/transl_table=11
/product="3-deoxy-D-arabinohexulosonate-7-phosphate
synthase"
/protein_id="AA086599.1"
/db_xref="GI:21959964"
/translation="MNYONDRLRIKEIKELPPVALLERFPSTVNAETVAETRNAIH
NILAQRDRLIVYPCSIHDTQAKYARLARLAREELQGELEVVRVYEFKPTTV
GWKGLINDPHMNSYDINEGILARELIVEINSGLPAGEFLDITTPYLADMSWG
AIGARTTESOVHRELASGLSCPVGFKGTGTGTTKVAIDAINASAPHCSTKMGHS
AIVTAGNCDCCHILLRGKEPNSSKRVNVAKGLIKAGLEPQIMDPSHANSKFO
KOMEVGTVCQOIAQGEKSIIGVIESHVEGSONLESGBLTYGOSTIDACIGMDT
EALLROLAGVARRRQ"
3294..4232
/gene="y3050"
CDS
3294..4232
/gene="y3050"
/function="putative transport"
/note="residues 4 to 310 of 312 are 66.12 pct identical to
residues 9 to 313 of 313 from E. coli K12 : B0752;
residues 9 to 309 of 312 are 67.44 pct identical to
residues 14 to 312 of 312 from Genpept : >gb|A019697.1|
(AE008731) putative CDF family transport protein
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="putative transport system permease protein"
/protein_id="AA086600.1"
/db_xref="GI:21959965"
/translation="MAVSTFFSDSNSKRLIAFAITTLFVTEAIGWLSGSLALLA
DAGHMLTDSALFIALMAVHFSORRDPDPRFTGYRLTTLAFAVNAALLIVLITW
EAVHRTSPHEWAGCTPMLITAILGLANIEFCWILHKGEEKINVRAAALHVSDDL
GSVGAMTAAIVILITGMTPIIDPLISYIVLILRSWRLKESFHELEGAPOEDIN
KLKRDICTNITYEVNRIHVMGVGRMLTLHDOYTPPLDALLQRIQDYLLHNYR
ISHATVOMEYOHCGTDPDGCINQAPADGHHRRHHE"
complement(4347..5072)
/gene="pnuC"
/note="y3051"
complement(4347..5072)
CDS
/gene="pnuC"
/function="transport; transport of small molecules;
Nucleosides, purines, pyrimidines"
/note="residues 1 to 238 of 241 are 76.05 pct identical to
residues 1 to 238 of 239 from E. coli K12 : B0751;
residues 1 to 241 of 241 are 100.00 pct identical to
residues 1 to 241 of 241 from Genpept : >emb|CAC6997.1|
(AJ414146) integral membrane NMN transport protein pnuC
[Verstina pestis]"
/codon_start=1
/transl_table=11
/product="nucleoside/purine/pyrimidine transporter"
/protein_id="AA086601.1"
/db_xref="GI:21959966"
/translation="MDRLSGNINIVHPIAGGYDLSMIEAIGTLGTLCTWPAKRE
ITNLYGLINLTVLAVITFOQLYASLLDLPFGCAITGVYAMSKOTPNQELKIR
WLSLPRALAMAVACIAGIVLMTLHIDIVFAMLTIVATNMGAGANOMPELOPDEP
FWDSTMYVLSIAAMITMTKRYEWNMLIWDIVLVAIFATQGYAMALEYAPLITLIL
LNGSWMTIKSARNSRPLSTOG"
complement(5194..6255)
/gene="nada"
/note="y3052"
complement(5194..6255)
CDS
/gene="nada"
/function="enzyme; biosynthesis of cofactors, carriers:
pyridine nucleotide"
/note="residues 1 to 346 of 353 are 82.08 pct identical to
residues 1 to 346 of 347 from E. coli K12 : B0750;
residues 1 to 346 of 353 are 82.94 pct identical to
residues 1 to 346 of 347 from Genpept : >gb|A019695.1|
(AE008731) quinolinate synthetase, A protein [Salmonella
typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="quinolinate synthetase"
/protein_id="AA086602.1"
/db_xref="GI:21959967"
/translation="KSEITFDVNAALITPPARVPDITNEKAFYREKIKTLKORDAVL
VAHYTDPPEIOALAEETGCVADSELMARFGNNHASTILVAGVFMKEPTKINPEK
KVIMPTLNAECSDLDGCPVDEFTAFCDSDPDFTVVAVNTSAVAKADWVYTSIAY
ELEIHDLSGEKLTMAPDRHLGSYVGRKSGAVLVCMOGACIVHDEFKIQALARKALY
PDAEIVLHPSPQAVVMDMDAVGSTSOLIOAKTLPOKTLIVATDRGIFYKMOQACPD
KELEAPTAEGCATCSGACHPMANNGRAIAEGLEGGVHHEIHDVEELROQALIP
LNRMLDANOLKLVQKNGA"
complement(6622..6691)
/product="tRNA-OTHEK"
/note="anticodon: TTT"
/function="tRNA: tRNA"
complement(6725..6797)
tRNA
tRNA
tRNA

```


Db	152652	-----ACAGAGAACCGCAGCAACCTGTCTTCTGTAAGACGACACGT	152693
Qy	61	AlaLeuAlaSerLysLeuPro-----SerLeuValTyrPheAspPheAspSerAspGlu	78
Db	152694	CTTCAGATGCAGACGCTCAAAAGAACACATGCTTACTCTGCTTCCGTTCAATATACGAT	152753
Qy	79	IleLysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrAsnGln	98
Db	152754	ATCGGTTCTGACTTCCTCCATAAAGCTGGATGACACAGCTGCATTCCTCGGTACCAACCA	152813
Qy	99	ThrAlaArgValLeuValAlaGlyLysThrAspGluArgGlySerArgGluTyrAsnMet	118
Db	152814	TCTGACAAGATTGTTTCTTGAAGCCACGCGGATGAACGGGGATGCGCAATAACCAACAC	152873
Qy	119	SerLeuGlyGluArgArgAlaValAlaValaArgAsnTyrLeuLeuGlyLysGlyIleAsn	138
Db	152874	GCTTTGGGAGAGCGTGTCTGACGACAGTGAATAATGTACTGCAGAGTAAAGCGTTTCT	152933
Qy	139	GlnAlaSerValGluIleLeuSerPheGlyGluGluArgProIleAlaPheGlyThrAsn	158
Db	152934	GCTGACACAGATCTCTATCGTTCTCTTACGGTAAAGAGAAACACAGCATGACTGGCGCATGC	152993
Qy	159	GluGluAlaTyrPserGlnAsnArgArgAlaGluLeuSerTyr	172
Db	152994	GAGACGACGCTTGTGCTAAACCGTGTGTGACGTTTGTTC	153035
RESULT 22			
AE012414/c			
LOCUS	AE012414	12393 bp	DNA
DEFINITION	Xanthomonas campestris pv. campestris str. ATCC 33913, section 322		
ACCESSION	AE012414		AE008922
VERSION	AE012414.1		GI:21114210
KEYWORDS			
SOURCE			
ORGANISM	Xanthomonas campestris pv. campestris str. ATCC 33913. Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas.		
REFERENCE	1 (bases 1 to 12393)		
AUTHORS	da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteleiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.E., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Canavan, F., Cardoso, J., Chambergo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kish, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, V.F., Locali, E.C., Machado, M.A., Madella, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandri, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Seta, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tanura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.		
TITLE	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities		
JOURNAL	Nature 417 (6887), 459-463 (2002)		
MEDLINE	22022145		
PUBMED	12024217		
REFERENCE	2 (bases 1 to 12393)		
AUTHORS	da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,		

TITLE	
JOURNAL	
FEATURES	
source	
Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spínola,L.A.F., Taktá,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P. Direct Submission Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil	
gene	Location/Qualifiers
CDS	1..12393
	/organism="Xanthomonas campestris pv. campestris str. ATCC 33913"
	/strain="ATCC_33913"
	/db_xref="ATCC:33913"
	/db_xref="taxon:190485"
	/note="pathovar: campestris"
	88..291
	/gene="XCc2999"
	88..291
	/gene="XCc2999"
	/note="putative; ORF located using Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AA042271.1"
	/db_xref="GI:21114211"
	/translation="MYVCDNPVCASYGHTLTGLCSPSGMPHANSDLPPPTGYLRA MAKAKREVAANAQMOLDPAAP" 470..595
	/gene="XCc3000"
	470..595
	/gene="XCc3000"
	/note="putative; ORF located using Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AA042272.1"
	/db_xref="GI:21114212"
	/translation="MAQLMLRGLMSSRTSHMALATDAMEILDYGSAREISA" 532..831
	/gene="XCc3001"
	532..831
	/gene="XCc3001"
	/note="putative; ORF located using Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AA042273.1"
	/db_xref="GI:21114213"
	/translation="MSOSNGMATQADRPVDASQAQQOYVAPHKKREQADYLRLVE AHLASGAEEVTTPRIPLKSIALALNRSTVGK" 859..352
	/gene="orf37"
	859..352
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="phage-related protein"
	/protein_id="AA042274.1"
	/db_xref="GI:21114214"
	/translation="MLQRLEDDYGLKHSRGTSYMRGVCPACSKKETLPFPKPVIK CGREKAGHELHVLDYDLDFDDMSKRPMQASPRAADAYLBESSRFALAPLRGLT TOESYDIKKKEGTAVTFEALDKGWMRRLLDRPHRRKQARHPGCSVAGWMCAP AAELMRATEWTVIEGITFDAILDQHGCVASMSCAFDESLRLDAKLRAQNLP LVKGIDNPGARDYTHKHARRADALGENSRALLAQPVTKRIDMNIDLPRQAQGS QKWMDALTEARYGODLMARSATEKGLMTDNQASDFMLEYSRLYWPEFDTRFE PFRHNSGKFTFGTGHISASEFKRKLSLAAGAMTGSGHQDLRLIEBOEAIKTYE AIDVVGSKERRAVILGDIAVRGEVYTANEDYFSRKKLALKSTQSIRLEIORDPE AFRRDMLEPWLMQCETGHTGMAMTWFSLSAEQIRAGHKSPPLEATGEGAGCTTLT


```

/codon_start=1
/transl_table=11
/product="phosphoserine aminotransferase"
/protein_id="BAB50673.1"
/db_xref="GI:14024069"
/translation="MTTHRRPGGLRPANPNFSSGCPAKKRPQWSVEALAKNALGSRSHAK
IGTKRLEOALREILQVPAHRIGIVPASPDAVEAMALMSLGERGVDMAMESEFG
SGWVDVVKOLKLPDVKRLKLEAGYGLDLAKIDEDRDVVTWNGTSGVRPNDEIP
ADKKGLTICDATSAAFAORLEDEKLDVTFMSQKVLGEGAGHGMILISPRAVARLESY
KPMPLPKIERLTSGGKLEIGIKPEGIINPSMLCEYEDLDALHMAKSGIGLDIVAR
ADANAVALDEGVSSSMGLHAYQAPATRSNTSYCSFTPPDVAALDADQATFAKGLV
SALDKGVAVDGVSSYSDMGPGLKINMGCAIYVTSDELTLPLDMAPAAKASLKAAA"
complement(3096..3869)
/gene="m113877"
complement(3096..3869)
/gene="m113877"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50674.1"
/db_xref="GI:14024070"
/translation="MTLSRIYALALAGLILPLTPAVADYDPPIYADQAPPVPEV
GSGWYLRGVSYLAOKSFKNDDPAFTPAFSEDEKEDPIFASIGFCHENDYLEADNLG
VGLGNKIGIGYDLSISVPATSVASADLKAYSLMNAIVDGLTGYITPYLGGG
VGVOSTRRLSANFTPNNGDPTDDEVDQDKTKYSLAYTLNGLAYQVSKNSVDLGY
QYFSAPDAEYVNASLTSFPHKSGISNHQIKLGRYDLA"
complement(4111..4944)
/gene="m113878"
complement(4111..4944)
/gene="m113878"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50675.1"
/db_xref="GI:14024071"
/translation="MNTARKALLALIFAGLASPYPAADLPEPVEEAPPPYEDQPV
VGVYLRGIDYHKSVDRCIDMTYTGIDPCNCTVAPGKSPDYGLKGGFSLGGGVGY
KINDRLRAVYADIVFMSKSNFNGSTDVSTSEVSKMSALLLANAYVDIGTWHGITP
IYVAGIGGARVWMDYVDPTAETNPASNMRFAYALMAGASYCLDKIILDAGRFS
HIOGGRMFEMDASSAGPFDRGINTHEVAGLRYQFGNGGCAVAVAYQPEPEIYR
K"
complement(5193..6545)
/gene="m113879"
complement(5193..6545)
/gene="m113879"
/codon_start=1
/transl_table=11
/product="phosphoglucosyltransferase/phosphomannomutase family
protein: Msa"
/protein_id="BAB50676.1"
/db_xref="GI:14024072"
/translation="MAGNYFGTDGIRGRANKFPMTAETAMRVMAAGLSFGGSHRR
VVLGKDTLRSGMIENAMVAGIACAGMDVFLGPIPTPAVAMLVRSLSADIGVMISAS
HNPYDNGIKLEFGPDGYKLSDEIERIESMDKDELTLASDGLGRKRDYCHDRI
IEFAKRTLPKRSMSLSGLRIIVDCANGAAYKVAPEPLMELGAEVAINIEPGEINKE
CGSTHPAGLOKKVHEVRADIGIALDGDADRYIVDENCAIVDGDQIMAMIAESWMSQK
RLAGGCVSYVMSNLGLERFLGDKLQIHRKRVGDRVYEHMRHAGLVNGGEGSHIY
LSDFSTGDSVLSALQVLAICIKRQGRPVSELSKFEPPOLIKNVRILNGKPLEEAPY
KAAIEDARNRLKAGRLVIRPSGTEPLIRVMAEGDDQLVAVVNDIVEVISSETRSA
"
6913..7779
/gene="m113880"
6913..7779
/gene="m113880"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50677.1"
/db_xref="GI:14024073"
/translation="MLHRLATLGLAVVTLFLAANALTMACSTRYNAXKADGTL
AGQTNQFRKAQCGTDAATDTKPAITKAETPAKATETKPPAKAAAADSTAKGL
TSKESAKYQAQAKAGTNGKMDFRTECAAGASAAADTTKPAITKAAP
AGGKGLTMAECSAKYQAQAKTANTLKQGMKMFRAECGAAAADDSVPAPSEANYTKE
"

```

```

PAKPTTVA PKGTFPSAISPKYASFTPGKGMHTCLDQYALKDNALGLKWTQKG
GFSYLCNAKKG"
8079..8537
/gene="m113881"
8079..8537
/gene="m113881"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50678.1"
/db_xref="GI:14024074"
/translation="MPNLKLAIVLAVTASVANSVAPASALTMKEGGEYVYRTAKEGTIN
GMDMAFERKECAAPPAESVGDYVTKTPQKRTKDAVAPVAPVAPVPTTLAAEF
KTEKAKARKMCTQDGYNGNEACTLGNLWTKQGGTYSICNRLAES"
complement(8624..10552)
/gene="m113882"
complement(8624..10552)
/gene="m113882"
/codon_start=1
/transl_table=11
/product="metalloprotease (cell division protein): FtsH"
/protein_id="BAB50679.1"
/db_xref="GI:14024075"
/translation="MNPYRMLALMAIIVALLALFNLFPOTRGASDPVPSOFIO
DVAAGRYVKTITAGARITGYTQNSGTGEPYSPDQVLSRDKNTTINARPEADS
NSLGGYLSMTPLIIGVAFEPFRQSSGSRAMGFSKAKLITAHGRVTFPDA
GVDAKEDLEETVEFLDPQKFORLQKIPRGVLLVGPFGTKTLARSAVAGANPE
FTISGSDVEVFVGVASRVMDPDKKNAKPCITFIDEIDAVGRHGAAGLGGNDER
EQLNOLULEMDGFESNESIILIAATNRPLVDLALRPERDQVVPVPPDIYGERK
ILKHAVRNVPPLAVNDELKVVARGTPEGSGADLMVLVESAALMAARRKRLVTMAEED
AKDKIMGAERSSAMTOAKELTAYMGHAGHATLALNPSADPLHRTITIPRRAIGM
VTNMGFSDKLCHVAVGNQEBVFLGHSVARTONISETAQIIDAVERNLIIDEAVSTAK
SLTKKKKEWTAQGLLEYFTTSGDEIKOLINGEKPARDLGDTPSRGSAVPKSSG
RRKKPEEGKEWAPQPS"
complement(10675..12039)
/gene="m113884"
complement(10675..12039)
/gene="m113884"
/note="Similar to yaen protein [Bradyrhizobium japonicum],
contains similarity to cell cycle protein Mesd/cytosine
deaminase-related protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50680.1"
/db_xref="GI:14024076"
/translation="MLDTEPDLSTRFLSHIDTFHGAVAAYSGSDSTALLLKHNF
RTGSSAKLAVITIDHGLRGSAEAQVAKICAERGIAHRTITWTRKPSGGLPAAR
DARYHLIAEARAAGIGLIVTGHADQAEVLMRHADEIADLAGRIGAGMAPATL
YMRWEMIVRELGTGRSALRDLPRREHVGAEEDPTINIDEAEPFRBALAGSAGHM
ENTLRLAGOAVERGOLGASANLIRIASQEPSTGLRIDPALLIDDOAATIALKIL
LATRAGAAFTPDARSALNVRILKAGLCTLSRVVDRFHAGLFLRRARGLPPATL
AVDNTLIDGRRHITLINDMSGALLIAPGAAARLAIIDGEPASILIRALLAEPIL
QAFENLITPGDEPRLPVFAARPVVSPFARLPFSDIAPARAVALIGASPLPASPLGG
HSD"
complement(12098..13210)
/gene="m113886"
complement(12098..13210)
/gene="m113886"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50681.1"
/db_xref="GI:14024077"
/translation="MLIRSVLSTLAILLSGFTALASGPGASGPGQSTDSGFSPHLP
SLEPLSLFGDKKPPDOVQFPAQDSNGAGLEDOQLRQNNKGLIEENFQILQOQMRK
QOEDNEFERQOLEGAGQGOAPPAPKSKDATTTNTDVAAPATQAPADAGAGGNOS
AGKTEVEDIVESPEGDGKVIPTGAPKGTFTSIVDNGNVITDAGNGVOSTAPTOP
AAPATKAPAKAGKSNCTIVAAIPSTNDPELVKNSYQFLLSDGYTAEGFPHDSIRF
"

```

Alignment Scores:
Pred. No.:
Score:

4.45e-16 Length: 349116
269.50 Matches: 64

Percent Similarity: 55.97% Conservative: 25
 Best Local Similarity: 40.25% Mismatches: 53
 Query Match: 31.30% Indels: 17
 DB: 1 Gaps: 5

US-09-674-779b-2 (1-172) x AF003003 (1-349116)

QY 13 AlaleuSerValleuThrheMetThrGlyCysAlaAsnLysSerThrSerGlnValMet 32
 DB 13885 GCGCTGGTGGCGGATGCTCCGACGCGCGGTTCGCTCGAAGAAGCG----- 13838
 QY 33 ValAlaProAsn-----AlaProThrGlyTyrThrGlyValIleThrGlyValAla 50
 DB 13837 -----CCGACACACCGCCCGCCGACCTCGGCTCAACGCG-----GCCGGTGGCGGA 13793
 QY 51 ---ProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSerLeu 69
 DB 13792 ACCCGCGGCTGGCGGACGACGCTTCAACCGTC-----AACATTTGGCGACCGCC 13748
 QY 70 ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGlu 89
 DB 13747 ATCTTCTGATACGACGCTGCTCTGATCCGCGCGCGCGACCGACCGACCTCGCTGCC 13688
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
 DB 13687 CAGCGCGACGCTGTAACCGACGACGACGACGACGACGACGACGACGACGACGACGAC 13628
 QY 110 GluArgGlySerArgGlyTyrAsnMetSerLeuGlyGluArgGlyAlaValAlaValArg 129
 DB 13627 GAGCGCGCGACGCGCGGATCAATCAATCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 13568
 QY 130 AsnTyrLeuLeuGlyGlyGlyIleAsnGlnAlaSerValGluIleLeuSerPheGlyGlu 149
 DB 13567 GACTTCTGCTGTTCCAAAGGTGTCGCTGCGCGCGCGCTGAAGACATTTCAATCAATG 13508
 QY 150 GluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgAla 168
 DB 13507 GAACGCTCGGTCGCGGTCTGCGACGACGACGACGACGACGACGACGACGACGACG 13451

RESULT 25
 AEO04009/c 10833 bp DNA linear BCT 15-JUN-2001
 LOCUS Xylella fastidiosa 9a5c, section 155 of 229 of the complete genome.
 DEFINITION AEO04009 AE003849
 ACCESSION AE004009.1 GI:9106980
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Xylella fastidiosa 9a5c.
 Xylella fastidiosa 9a5c.
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 Xylella.

REFERENCE
 AUTHORS
 1 (bases 1 to 10833)
 Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
 Alvarange,R., Alves,L.M., Araya,J.E., Bala,G.S., Baptista,C.S.,
 Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
 Bueno,M.R., Camargo,A.A., Camargo,L.E., Carriro,D.M., Carier,H.,
 Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
 Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorriy,H.,
 Facinani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
 Fraga,J.S., Franca,S.C., Franco,M.C., Fromme,M., Furlan,L.R.,
 Ganier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
 Ho,P.L., Honeisel,D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.,
 and Marino,C.L.
 The genome sequence of the plant pathogen Xylella fastidiosa. The
 Xylella fastidiosa Consortium of the Organization for Nucleotide
 Sequencing and Analysis
 Nature 406 (6792), 151-157 (2000)

JOURNAL
 MEDLINE
 PUBMED
 10910347
 2 (bases 1 to 10833)
 Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
 Alvarange,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S.,
 Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
 Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carriro,D.M.,

TITLE
 JOURNAL
 FEATURES
 source
 1..10833
 /organism="Xylella fastidiosa 9a5c"
 /db_xref="taxon:160492"
 /clone="9a5c"
 192..2081
 /gene="Xf1888"
 192..2081
 /note="similar to SP1P45740 (percent identity: 67 %/query alignment coverage: 91.4 %/subject alignment coverage: 97.5 %) identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="chiamine biosynthesis protein"
 /protein_id="AAF84694.1"
 /db_xref="GI:9106981"
 /translation="MDRAARPPMORRRRRPPPSISGRTVITTAALPLPPLILSE
 KYTAPLIGARKIYITGSRQDIPREIRELALIPSSARRGNGENSLAYDNGITDPO
 ATIDLACGLPRLRAMIDERADYEAALYFVPSPVSTAYTAPFPAPRRARNAV
 TOLEYARGLVTPMEMEVALIREOQRRQEVNLRGORGADGAWALGVPTTPEVRD
 EARGRALPNNINHPSEPMIIGNFPLKINAINIGTALSSTIAEVEKIVAIIRG
 ADTINDISTGNDIHATREMIIRNSPVPGVPIYQALEKVGSHVEAKVAFDITLIE
 QAEQGVYVYTHAGVLRPFITPLPASRLIGTYSRGSSTMARWCQAHRSNPLVTFEEL
 CEIMAYAVAFSLGSLRPGCTADANDAGRAELRLEITGLTHIMNNQVYMGVPGH
 VPMHITKANMKOLACGAEFPYTLGTLTIDYAGYDHTISATGAMMGITAMLT
 HTSPKRLALPNLDVHDHDIIVAKIHAADALAKHPAQAQDADLSAREFERMDQF
 HSLPEKALALHDESLPKEAHRAFCSMGQPCFMSKISQEVADSSNLSGNTN
 TNDAL"

gene
 CDS
 2213..2992
 /gene="Xf1869"
 2213..2992
 /note="similar to SP1Q44532 (percent identity: 59 %/query alignment coverage: 97.3 %/subject alignment coverage: 97.7 %) identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="ferredoxin-NADP reductase"
 /protein_id="AAF84695.1"
 /db_xref="GI:9106982"
 /translation="WSPAFGTETVYVHHMTDAYSFITTDGSGRPNQGVNIGLE
 TPARKLRYVSLASNAWEQLEISIKVONGLSLRLKIRPKRLITGKRPSTLTLL
 HDLHGKRLYLIGTGLAPWMSIIKDEYTERFQVLTITGVRISKDLARPFKE
 LPQHEHLETIRKILLYPAVYREDFPPNRGRILHIDEGCANQNTGLPIIDQANDREM

gene
complement(2986. .3558)
/gene="XF1890"
CDS
complement(2986. .3558)
/gene="XF1890"
/note="Similar to SP1032770 (percent identity: 50 %/query alignment coverage: 82.6 %/subject alignment coverage: 100.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="glutathione peroxidase-like protein"
/protein_id="AAF84696.1"
/db_xref="GI:9106983"
/translation="MHTVSHAFASKLSLSDRPPMSITPTPTLDRGQALADRCQVLLVNASRCGETPOYAGLEMLQRYADAGLIVGFCDDPAGDEDEKILAEFCLLVYDPEPAKIKVNGADHPIMQWLRHRRGLFGMAIKMNTFKLIGNGQPIARSPKSPQLEHYIYALGEBGGNT"
3884. .5416
/gene="XF1891"
/db_xref="GI:9106984"
/note="Similar to SP1P94408 (percent identity: 35 %/query alignment coverage: 96.7 %/subject alignment coverage: 100.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="gi-cripeptide ABC transporter membrane protein"
/protein_id="AAF84697.1"
/db_xref="GI:9106984"
/translation="MSVDTTVKNTVYATPTQLPKFTTFGHPKPLMLMAEFWERFATYGRNALTYVTOFPDGNISGEENASTYGAVALYASAIFFGFAADRIQGRSILIGAVIMACGLITVPSRMEFELGATVYNGLEKFNISLIGOLYADGDRDRTIIFYMGINAGSLISPLTSMWAGOVGTQMOQYKVFASGGMISLMPVIGKRLKIGLEPPKDESIFFRFLIFGALAIPLAYILAKINATITAMTILGIFALALITIVATLRNGKIQDRVYALIIFFENFMWFFQAGSSSENFKNATVDSQILSNWEPVGMOSVNPPLAIIILAPILIVIMSILDKRIEPIKRSGLMGFNGFQILMVALSNLNASMTIPFSLIAVYIVQVGECLSPITGSMVTKLAPVAVGFAMGFWELSTAGNNLSGVFAFSEVSGEEMTVSSALRGYTFGWSLIGSILFLISPLINRHHGVR"
complement(5578. .6342)
/gene="XF1892"
CDS
complement(5578. .6342)
/gene="XF1892"
/note="Similar to SP1P32679 (percent identity: 53 %/query alignment coverage: 83.1 %/subject alignment coverage: 94.6 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="endonuclease V (deoxyinosine 3'endonuclease)"
/protein_id="AAF84698.1"
/db_xref="GI:9106985"
/translation="MPYALISDTGMVFMKISSIDIFAGMDGSTTEARLOSMAE RYLKDLNLSLEPTLAGDEGEDEGTAAAYIAMACDKILETVVYPTSPMYVYGLSFRELPAALQALOSRIPALVVDGGAHAPRLGIAAIFGLVTLNPGIVAKRIVGDEPEGTAFGEHTPIILHCTGVGMALRSKIRKPKMISPGKISLHSAITW TORCLNGYRLPEPTROADRLASRQKIVSLPSL"
6397. .7146
/gene="XF1893"
/db_xref="GI:9106986"
/note="Similar to SP1P31217 (percent identity: 56 %/query alignment coverage: 98.8 %/subject alignment coverage: 98.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phosphoglyceromutase"
/protein_id="AAF84699.1"
/db_xref="GI:9106986"
/translation="MTRKLVILRHGOSQWNSMNPRTGWDVIGLTFEQGHQDEATMAGHLM

KEEGLFEDVAHTSLKRAIHTLQDALKALQDMLPIYKSWRLNERRHAGALQGLDITDAKHEQOVNIMRSYDIOPPIDIDPSPHMDRRRAALDRKVLVPTSSIKTLERLPYWYDAIAPOLDNDKTVLSAHGNSIRAIYKXKILNDESDKILNVAIPIGILFELS DTLQVSYRYLGDPPDAQORASEWVANGKAK"
complement(7245. .7928)
/gene="XF1894"
CDS
complement(7245. .7928)
/gene="XF1894"
/note="Similar to GI15514780 (percent identity: 63 %/query alignment coverage: 93.0 %/subject alignment coverage: 98.1 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84700.1"
/db_xref="GI:9106987"
/translation="ANDQRNNRNLNATRSQKISIFLISLQGEANSVGMPTVYVRLTGCPRLRCYCDIATAFHGEWCSDITIVSEVRSYGVHVCVTGEGELAKRCLLEKLCDAQFEVSLSTSGALDIAVADVLRVSRVDIKTPGSEAHRRNHMPNLAITLPPDQIKFVLCRSRADVEMARTCVAEHELEERRCMVWFSPSKODIAPTVLADMIISRLGVRQDLHK LLMNDEGR"

Alignment Scores:

Pred. No.:	1,37e-17	Length:	10833
Score:	265.50	Matches:	60
Percent Similarity:	53.89%	Conservative:	30
Best Local Similarity:	35.93%	Mismatches:	54
Query Match:	30.84%	Indels:	23
DB:	1	Gaps:	5

US-09-674-779b-2 (1-172) x AE004009 (1-10833)

QY	20	MetThGlyCysAlaAsnlySerThrSerGlnValMetVal-----	33
DB	9250	CTTGTGCGCTGTTCAAAAAAGTTAAGAGCAACCAACGATGCGTCAAACTATGGCA	9191
QY	34	-----AlaPro-----AsnAlaProThGlyTyrThGlyValIle	45
DB	9190	CCACACTGTTTCAAGCCCCCGCCACCCACCTACGCGCCCAACGATTCCTTGGG--CTT	9134
QY	46	TyrThGlyValAlaProLeuValAspAsnAspGluThrValIleValAlaLeuAlaSer	65
DB	9133	TACACA-----CGCGCTACCTAGTACCACTGCTTGGCT-----	9095
QY	66	LeuProSerLeuValTyrThrAspHisAspSerAspGluIleLysProGlnAlaAla	85
DB	9094	---CAACGTGCTGCTATTTCACCTTGTATTAAGATGATGAGAAAGAAATTCMAACG	9038
QY	86	IleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnIleThrAlaArgValLeuVal	105
DB	9037	GTTCTTGGTGGTCAATGCAAAATATCTTGGCACACCTCCCTGACACATTCACG	8978
QY	106	GlyHisThrAspGluArgGlySerArgIleLysAsnMetSerLeuGlyGluArgAla	125
DB	8977	GGCAATACGATGAGACCGCGGTGCGGTGAGATATATAGCTCGGAGAACCGCGGT	8918
QY	126	ValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIle	145
DB	8917	AACTCTGTGTGTATTCATTACAGCGAATGCTCTGCGGACAGATTGAACGTTGTT	8858
QY	146	SerPheGlyGluGluArgProIleAlaPheGlyThrAsnGlnGluAlaIleArgSerGlu	165
DB	8857	AGTTACGCTGAAGAGCGGTGCGGTGACCTGATCAATCAACAGAAATGCTGTACGTAC	8798
QY	166	ArgArgAlaGluLeuSerTyr	172
DB	8797	CGGCGCTTGAATCTCTAT	8777

RESULT 26
AR089419 462 bp DNA linear PAT 07-SEP-2000
LOCUS
DEFINITION Sequence 178 from patent US 5994066.

ACCESSION	AR089419	GI:10016176
VERSION	AR089419.1	
KEYWORDS		
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	Unclassified.	
AUTHORS	1 (bases 1 to 462)	
TITLE	Begeon,M.G., Picard,F.J., Ouellette,M. and Roy,P.H.	
	Species-specific and universal DNA probes and amplification primers	
	to rapidly detect and identify common bacterial pathogens and	
	associated antibiotic resistance genes from clinical specimens for	
	routine diagnosis in microbiology laboratories	
	Patent: US 599406-A 178 30-NOV-1999;	
JOURNAL	Location/Qualifiers	
FEATURES	1..462	
source	/organism="unknown"	
BASE COUNT	142 a 85 c 103 g 132 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	3.47e-19	Length: 462
Score:	264.00	Matches: 51
Percent Similarity:	68.93%	Conservative: 20
Best Local Similarity:	49.51%	Mismatches: 32
Query Match:	30.66%	Indels: 0
DB:	6	Gaps: 0
US-09-674-779B-2 (1-172) x AR089419 (1-462)		
QY	70 ValTYRPhenAPhAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGlu 89	
Db	151 GTATATTTTGGTTTGATTAATACGACATCCACCGGTGAATCGTTCAATCTTAGATCG 210	
QY	90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109	
Db	211 CACGACAGCATATTTTAATGCACACGCGAGCTGCTAAAGTATGTAGTAAGGTAACTGAT 270	
QY	110 GluArgGlySerArgGluTyrAsnMetSerLeuGluArgAlaValAlaValArg 129	
Db	271 GACCTGGTACACGCAATACAACTGGATTAGGACCAACGTCGTGCAGATCACTTAA 330	
QY	130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149	
Db	331 GGTATTATTGACAGTAAAGTGTGATGCTGGTAATTTGGCACAAGTATCTTACGCTGA 390	
QY	150 GluArgProIleAlaPheGlyThrAsnGlnGluAlaTyrSerGlnAsnArgArgAlaGlu 169	
Db	391 GAAAAACCTGCAGTATTAGTCACGATGCAATGCAATTTCTTAAAAACCGTCGTGCAGTG 450	
QY	170 LeuSerTyr 172	
Db	451 TTAGCGTAC 459	
RESULT 27		
LOCUS	HEA15KLP	737 bp DNA linear BCT 26-APR-199
DEFINITION	H.influenzae PC protein (15kd peptidoglycan-associated outer	
ACCESSION	membrane lipoprotein), complete cds.	
VERSION	M18878.1	GI:148859
KEYWORDS	lipoprotein; outer membrane lipoprotein.	
SOURCE	H.influenzae (strain 52, sub-species Rb) DNA.	
ORGANISM	Haemophilus influenzae	
	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;	
	Haemophilus.	
REFERENCE	1 (bases 1 to 737)	
AUTHORS	Deich,R.A., Metcalf,B.J., Finn,C.W., Farley,J.E. and Green,B.A.	
TITLE	Cloning of genes encoding a 15,000-dalton peptidoglycan-associated	
	outer membrane lipoprotein and an antigenically related	
	15,000-dalton protein from Haemophilus influenzae	
JOURNAL	J Bacteriol. 170 (2), 489-498 (1988)	
MEDLINE	86115138	
PUBMED	2828309	

COMMENT	Submitted in computer readable form by R.A. Delch 06-APR-1988 This gene encodes a 15 kdalton lipoprotein associated with the peptidoglycan layer. Predicted amino acid sequence is identical to that of the H. influenzae PC protein (Nelson et al., Infect Immun 56, 128-134, 1980). Protein sequence is also 608 homologous to the E. coli pal gene product (Chem and Henning, Eur. J. Biochem. 163, 73-77, 1988).			
FEATURES	Location/Qualifiers			
SOURCE	1..737			
	/organism="Haemophilus influenzae"			
	/db_xref="taxon:727"			
CDS	241..702			
	/note="PC protein precursor"			
	/codon_start=1			
	/transl_table=11			
	/protein_id="AAA2940.1"			
	/db_xref="GI:148860"			
	/translation="MNKFKVSLYAGSVAAALACSSNNDAAGAAOTFGYSVDLQQRNTVYFGFDKVDITGERVQILDPAATLNTPAKTVKGNTPGRPEVNIALGQARDVAKYGLAGVDAGKLGTVSGEEPALVGDEAVSKNRRAVLAY"			
	241..297			
sig_peptide	/note="PC protein signal peptide"			
	298..699			
mat_peptide	/product="PC protein"			
	224 a 134 c 155 g 224 t			
BASE COUNT	1 bp upstream of BglII site.			
ORIGIN	1			
Alignment Scores:				
Pred. No.:	6.32e-19	Length:	737	
Score:	264.00	Matches:	51	
Percent Similarity:	68.93%	Conservative:	20	
Best Local Similarity:	49.51%	Mismatches:	32	
Query Match:	30.66%	Indels:	0	
DB:	1	Gaps:	0	
US-09-674-779B-2 (1-172) x HEA15KDP (1-737)				
QY	70	VALTYRPEAHPHESPASERASPGLULELPS	PROGINALALALALELEUASPGLU	89
DB	391	GTTTATTTGGTTTGATTAATATGACATTACTGCGTAATACGTTCAATCTTGATGAGCG		450
QY	90	GINALAGINPHELEUTHRRASNGINTHRALARGVALLEUVALALAGLYHISTHRASP		109
DB	451	CACGCGCATATTTAATGCAACACACAGCTGCTAAGATTAAGAGAAGGTAACATCGAT		510
QY	110	GLUARGGLYSEARARGLUTYRRASNETSERLEUGLYGUARGARGALAVALALVALARG		129
DB	511	GACGCGGTACACCGAATACACATCGCATTAAGCCACACCGTCGCGAGATGTAA		570
QY	130	ASNTYRLEULENGGILYLSGLYILEASNGINALASERVALGIULLEISERFHEGLYLU		149
DB	571	GGTTATTATTCGCTGAAGTGTGATCGCTGTAATTAAGGCACAGTATCTTACGCTAA		630
QY	150	GLUARGPROTILEALAPHEGLYTHRRASNGIUGLUALATPISERGLASARGALAGLU		169
DB	631	GAATAACCGCAGATTAAGTCATGATGAAGCTGCATATTTCAATAAACCGTCGTCAGTG		690
QY	170	LEUSERTYR		172
DB	691	TTAGCGTAC		699
RESULT 28				
LOCUS	106718	737 bp	DNA	linear
DEFINITION	Sequence 3 from Patent WO 9002557.			
ACCESSION	106718			
VERSION	106718.1	GI:589591		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unclassified.			
REFERENCE	1 (bases 1 to 737)			
AUTHORS	Antionis,A., Seid,R.C.J., Delch,R.A., Zlotnick,G.W. and Green,B.A.			

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

QY 70 ValTyrPheaspPheaspSeraspGluIleIysProGlnAlaAlaIleLeuaspGlu 89
 DB 218 GTATTATTGGTTGATTAATACGACATACCGGATGACGTTCAATCTTAATACGCG 277
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGluHisThrAsp 109
 DB 278 CACGACAGCATTTTAATGACACGACGCTGCTAAAGTTTACTAGACAGTAATACTGAT 337
 QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 DB 338 GAACGGGTATACACAGATATACACATCGCTTAGACACACGTCGACAGTACAGTAA 397
 QY 130 AsnTyrLeuLeuGlyIysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
 DB 398 GGTATTATTACAGTAAAGCTGTGATGCTGTAATATAGGCACAGTATCTTACGGTGA 457
 QY 150 GluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgAlaGlu 169
 DB 458 GAAACCTGCATATTAGTTCACGATGAAGCTGCAATTTCTTAACACGCTGTCAGTG 517
 QY 170 LeuSerTyr 172
 DB 518 TTACGCTAC 526

RESULT 31
 LOCUS 105561 867 bp DNA linear PAT 02-DEC-1994
 DEFINITION Sequence 1 from Patent EP 0281673.
 ACCESSION 105561
 VERSION 105561.1 GI:590761
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 867)
 AUTHORS Murphy, T.F. and Apicella, M.A.
 TITLE Plasmid for production of membrane protein, bacterium containing same, monoclonal antibody therefore, and method for the identification of haemophilus influenzae
 JOURNAL Patent: EP 0281673-A1 14-SEP-1988;
 FEATURES
 source location/Qualifiers
 1..867
 /organism="unknown"

BASE COUNT 253 a 167 c 175 g 271 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 7.78e-19 Length: 867
 Score: 264.00 Matches: 51
 Percent Similarity: 68.93% Conservative: 20
 Best Local Similarity: 49.51% Mismatches: 32
 Query Match: 30.66% Indels: 0
 DB: 6 Gaps: 0

US-09-674-779b-2 (1-172) x 105561 (1-867)

QY 70 ValTyrPheaspPheaspSeraspGluIleIysProGlnAlaAlaIleLeuaspGlu 89
 DB 218 GTATTATTGGTTGATTAATACGACATACCGGATGACGTTCAATCTTAATACGCG 277
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGluHisThrAsp 109
 DB 278 CACGACAGCATTTTAATGACACGACGCTGCTAAAGTTTACTAGACAGTAATACTGAT 337
 QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 DB 338 GAACGGGTATACACAGATATACACATCGCTTAGACACACGTCGACAGTACAGTAA 397
 QY 130 AsnTyrLeuLeuGlyIysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
 DB 398 GGTATTATTACAGTAAAGCTGTGATGCTGTAATATAGGCACAGTATCTTACGGTGA 457

QY 150 GluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgAlaGlu 169
 DB 458 GAAACCTGCATATTAGTTCACGATGAAGCTGCAATTTCTTAACACGCTGTCAGTG 517

QY 170 LeuSerTyr 172
 DB 518 TTACGCTAC 526

RESULT 32
 LOCUS 13620 bp DNA linear BCT 29-MAY-1998
 DEFINITION Haemophilus influenzae Rd section 37 of 163 of the complete genome.
 ACCESSION U32722.L42023
 VERSION U32722.1 GI:1573348
 KEYWORDS
 SOURCE
 ORGANISM
 Haemophilus influenzae Rd.
 Haemophilus influenzae Rd.
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Haemophilus.
 1 (bases 1 to 13620)
 Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,
 Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,
 Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,
 Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,
 Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,
 Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M.,
 Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhman, J.L.,
 Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M.,
 Smith, H.O. and Venter, J.C.
 Whole-genome random sequencing and assembly of Haemophilus
 influenzae Rd
 Science 269 (5223), 496-512 (1995)
 95350630
 7542800

JOURNAL MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 2 (bases 1 to 13620)
 Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S.,
 Borodovsky, M., Rudd, K.E. and Koonin, E.V.
 Metabolism and evolution of Haemophilus influenzae deduced from a
 whole-genome comparison with Escherichia coli
 Curr. Biol. 6 (3), 279-291 (1996)
 96398784
 8805245

JOURNAL MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 3 (bases 1 to 13620)
 White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
 Direct Submission
 Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 4 (bases 1 to 13620)
 White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
 Direct Submission
 Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 The H. influenzae sequence has been updated by R. Fleischmann. New
 database matches have been assigned, product names have been
 improved, and a number of frame shifts have been corrected. We
 gratefully acknowledge the work of Tatusov et al. We have
 incorporated their annotation into the /notes fields of the
 corresponding H. influenzae genes
 5 (bases 1 to 13620)
 White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D.,
 Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.
 Direct Submission
 Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 The whole genome was shifted by 568 nucleotides for a new start
 On Sep 30, 1996 this sequence version replaced gi:1221050.
 location/Qualifiers
 1..13620
 /organism="Haemophilus influenzae Rd"
 /db_xref="taxon:71421"
 complement(61..513)
 /gene="HI0379"
 complement(61..513)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REMARK
 COMMENT
 FEATURES
 source
 gene
 CDS

```

/gene="HI0379"
/note="similar to GB:U00096 PID:1788880 percent identity:
63.38; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22037.1"
/db_xref="GI:1573349"
/translation="MKLSKGRVAVTAVIDALNADGPSVLADISERQHSIYLEQ
LPAKRDGLKSVKRGCGYQLGPSQISQVGMITAVNENIHVTCOLGKENCNV
ECITHMEDSLKIESLNEITLAEIVNKNRKRSHDFENLLVNO"
/complement(566..1291)
/gene="HI0380"
/complement(566..1291)
/gene="HI0380"
/note="similar to GB:U00096 PID:1788881 percent identity:
54.43; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22038.1"
/db_xref="GI:1573350"
/translation="MLENIRIVLLETSHSGNIGSAAMTGLTQLCLVSPSVDQ
SYLSAGAEIVKARVYDSFDEAVDCCPLVTSARLRLQNTLLEPRCAKVVAY
KGTATVGRERICLTNEELCKHYHINIPANDPYSILNAMAQVLSYELRAFLVQ
NNKNSLSLEKKNPTDOLAFPDYTERIYOSIGFTQONGVMKRLKLYYRAKLEKN
ELNLNGLMSAVEKRIDLTKEDN"
/complement(1405..1477)
/gene="tRNA-Lys-3"
/complement(1405..1477)
/gene="tRNA-Lys-3"
/gene="tRNA-Lys-3"
/product="tRNA-Lys"
/note="codon recognized: AAG"
/complement(1733..2194)
/gene="HI0381"
/complement(1733..2194)
/gene="HI0381"
/note="similar to SP:P10324 PID:1573351 SP:P10324
GB:M18678 PID:1573351 percent identity: 100.00; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptidoglycan-associated outer membrane
lipoprotein (pal)"
/protein_id="AAC22039.1"
/db_xref="GI:1573351"
/translation="MNKFKVSLVAGSVAALAACSSNNDAAGNAGAOPGCGVADL
QQRVNVYFGFDYDITGEVYQILDAAVYLNTPAKYLVESGNTDERGTPEPNTALG
ORADAVKGLACKGYDAGKTGVSGEERPAVLGHDEAYSKNRRAVLAY"
/complement(2218..3501)
/gene="HI0382"
/complement(2218..3501)
/gene="HI0382"
/note="similar to PID:1685080 percent identity: 96.96;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="colicin tolerance protein (tolB)"
/protein_id="AAC22040.1"
/db_xref="GI:1573352"
/translation="MKLIRKIVSFAIVAVGSNAFAGDEVIVIDEGYDGRPAIV
PFVGSAPEDISIKTVADDLRNSKRFNPINASQRPRTSAEYNPEAWNSIGIDAVIG
OVVPSNGSITYQLDITYGASGTGFTVIMONSYYTNKMLRYGAHVSDEVEFKLTA
IKGAFRTIRIAIVVQKNGSQPYEVRAVDYDGNQPIVNSAOPIMSPAMSPGQRLAY
VSENNKSQLVQDLSNGARKYVAFSGDNGAPAFSPDSRLAFASPGAGVINTIVMG
ANGGTPTQLTISGANNTEPAMPDGNLSILFTSDRSGSPDVPYMDASGSAVAVGGSG
AOSISADGKTLVINGNNVYKODLTGTGVEVSTGSLGSPSLSTNGIMITYSTQGL
GKVLQVADGGERKSLPGSDGVKFPAMSPVLR"
/complement(3543..4661)
/gene="HI0383"
/complement(3543..4661)
/gene="HI0383"
/note="similar to GB:M28232 SP:P19934 PID:148019 GB:U00096

```

```

PID:1651331 percent identity: 42.58; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="outer membrane integrity protein (tolA)"
/protein_id="AAC22041.1"
/db_xref="GI:1573353"
/translation="MQNNRQKGINAFISILHFTLFGLLISSLYHVEITMGGEQ
EGDIVAVIYDTGTAOEWGRIOQKGGADQKREPVEERKEPPNEETKHOEV
OOREELKROEOOROEIKCOEORALEOKOAEKAKQAQAAEAKLADAKR
LAAAKOAEPEAKAAEITAAQAKAEAKKALEAEAKAVAEKAAEAKAA
AAAKAKADAEAKATEAKRKADQASLDLDFLNGDGGGSASGNTNGGSGAA
LGGDGGKVGDDIAGVITKEIQRIQRLDFNGDGKVCRIQGRDGTILGYKISGD
DICSALSAVARKVPAAPSDIEYKSPILDEDIR"
/complement(4677..5096)
/gene="HI0384"
/complement(4677..5096)
/gene="HI0384"
/note="similar to SP:P05829 PID:1128980 GB:U00096
PID:1651330 PID:1786959 percent identity: 61.76;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="colicin transport protein (tolR)"
/protein_id="AAC22042.1"
/db_xref="GI:1573354"
/translation="MARQRKAIKSEINIVPLDVLVLIIFMATAPILISQVQEL
PDSVQSEVSNEDKVPVILEVAGIGKVAISIGSEOGGTEMTVQLSRQEDDMNT
LFLVGGAKEVPEEYVTKALNLHLAGIKSVGLMTNPI"
/complement(5168..5854)
/gene="HI0385"
/complement(5168..5854)
/gene="HI0385"
/note="similar to SP:P05828 PID:1128978 GB:U00096
PID:1651329 PID:1786958 percent identity: 68.47;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="colicin transport protein (tolO)"
/protein_id="AAC22043.1"
/db_xref="GI:1573355"
/translation="MTAELEFLDLFKASIVQVLYVILISFIIISNAIIORSRIIT
NALKPARTFEDRFSGEDLNKLESLNSRRQDGSQDIPFVGKRESRKQVAPDAP
EALIKGTBAMVLANNRITESLENNVPLAVASVSPYIGLGVWGMIAFMALSGA
KOATIQWAPGIAELIADTAIGLFAIPAVMAVYNLSLRVAIEDYGNFIDEPTTIL
HQAQGRKAPH"
/complement(5871..6281)
/gene="HI0386"
/complement(5871..6281)
/gene="HI0386"
/note="similar to SP:P08999 PID:1128977 GB:U00096
PID:1786957 percent identity: 53.08; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22044.1"
/db_xref="GI:1573356"
/translation="MLDNGFSEPVRYVEDTADAGVYHARYLHFEERARTEYLTLN
FTQOTLLEQDLAVVKTILADYCVAAKLDLILMVEVESEVKATILFPQRLMRNTL
MLSKATVACVADGKMKPVAFPREVAAAFPHLK"
/complement(6006..8528)
/gene="HI0387"
/complement(6006..8528)

```

Alignment Scores:
Pred. No.: 2,66e-17
Score: 264.00
Percent Similarity: 68.93%
Best Local Similarity: 49.51%
Query Match: 30.66%
DB: 1
Gaps: 0

Length: 13620
Matches: 51
Conservative: 20
Mismatch: 32
Indels: 0
Gaps: 0

gene		/translation="MAAETHPQAAHAAETHHQQAIKLYLWGLLPVLSALSYIDYFGAGGYLRSLILFIEMMLKAGLIIVFWMHAMERIALIYAIIIEPLLVIFVALMMS ESNVLTFRLLTPGGG"
CDS	5477..8038	
	/gene=".SMC01987"	
	5477..8038	
	/gene=".SMC01987"	
	/function="small molecule metabolism"	
	/note="Product confidence : putative"	

Alignment Scores:					
Pred. No.:	2,38e-15	Length:	340900		
Score:	262.50	Matches:	60		
Percent Similarity:	53.85%	Conservative:	24		
Best Local Similarity:	38.46%	Mismatches:	61		
Query Match:	30.49%	Indels:	11		
DB:	1	Gaps:	3		

US-09-674-779B-2 (1-172) x SME591791 (1-340900)					
Oy	13	AlAleuSerValleuthrPhemethnrglyCysAlaasnlySerThrserGlValmet	32		
Db	241535	GCGETCTCATMGACCGTGTGCCCTGTGGCTGGCTGCATGAAGAACA-----	241488		
Oy	33	ValAlaProXsnAlaProthrhglyTyrrhglyValIleTyrrhglyValAlaProLeu	52		
Db	241487	---CTGGCCAGCAATCGTCGCCGC---CTGGCTCTGGCGCAGCGCGCACCCCGGC	241436		
Oy	53	ValAspaSnaSpgluThrVallysAlaLeuAlaSerLyseuProserLeuValtyrPhe	72		
Db	241433	TTCGACGAGCACTTCACCGTC-----AACGTGGCGACCAGCATCTTCTC	241388		
Oy	73	AspPheaSPserAspEluileysProGlnAlaAlaIleLeuAspGluGlnAlaGln	92		
Db	241388	GATACGATTGCGATCGATCGATCGTGCCGACGCGACGACGCTGACCGCAAGCCGAG	241322		
Oy	93	PheIeuThrThrxsnGlnThrAlaArgValLeuAlaIlaGlyHisThrAspGluArgGly	112		
Db	241328	TGGCTGGCAAATAATCTCCGACTACGAAATCCATCCATGAGGCCATGCCGACCGCGGC	241269		
Oy	113	SerArgGluTyraMetserLeuGlyAlaArgAlaValAlaValaArgnsTyrrLeu	132		
Db	241268	ACCCGCGAATACAACTGGCGCTGGCGCGCCCGCTGTGGCGGACCGGACTATCTC	241205		
Oy	133	LeucglyLySGlyIleasnGlnAlaSerValGluIleIeserPheglyGluArgPro	152		
Db	241208	GTAGCGCGGCGCTCCCGGAAACCGCATGCGCACGATCTCTACGGGAMAAGAAACCG	241149		
Oy	153	IleAlaPheglyThrxsnGlnGluAlaTrpserGlnAsnArgAlaGala	168		
Db	241148	GTGCGCGTGGCAGCAGCATCTCCGTCGTCGTGTGGCAAGACCGCGCGCC	241101		

RESULT 34					
AE009302/c					
LOCUS	AE009302	10906 bp	DNA	linear	BCR 20-DEC-2001
DEFINITION	Agrobacterium tumefaciens str. C58 linear chromosome, section 72 of 187 of the complete sequence.				
VERSION	AE009302	AE008689			
ACCESSION	AE009302.1	GI:17742130			
KEYWORDS					
SOURCE ORGANISM	Agrobacterium tumefaciens str. C58 (U. Washington).				
REFERENCE	Agrobacterium tumefaciens str. C58 (U. Washington). Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium. 1 (bases 1 to 10906) Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutayavh, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saephammachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karpi, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Kreban, W., Perry, M., Gordon-Kamm, B., Rao, R., Xiao, S., Krause,				

TITLE	Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.
JOURNAL	The genome of the natural genetic engineer <i>Agrobacterium tumefaciens</i> C58
MEDLINE	Science 294 (5550), 2317-2323 (2001)
PUBMED	21608550
REFERENCE	11743193
AUTHORS	2 (bases 1 to 10906)
	Wood, D. W., Setubal, J. C., Kaul, R., Monks, D., Chen, L., Wood, G. E., Chen, Y., Woo, L., Kitajima, J. P., Okura, V. K., Almeida Jr, N. F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmeri, A., Raymond, C., Rouse, G., Seemhimmachak, C., Wu, Z., Gordon, D., Eisen, J. A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Bidle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.
TITLE	Direct Submission
JOURNAL	Submitted (127-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
FEATURES	98195-7242, USA
source	location/Qualifiers
	1. 10906
	/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"
	/strain="C58"
	/db_xref="taxon:180835"
	complement(113. .1489)
	/gene="mrSA"
	/note="Atu3709"
	complement(113. .1489)
	/gene="mrSA"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
	/codon_start=1
	/transl_table=11
	/product="phosphoglucose-1-phosphomannomutase"
	/protein_id="A144519.1"
	/db_xref="GI:17742131"
	/translacion="MPERGVQVMAARYFGTGDIGRQSNVFPMTPLAMRGVIAVGTIF RCHHRRVYIGKDTIRLGLMENAIVAGTAAAGDVLGPIPPVAMLRSLRAD IGVWISNHPFSDNGIKLFGPGYKLSIDELIEDLDKDIYQLAKRPEIGAKR VGDGIYRIEYFKRTLPRDVTLSGLRIATIDCANGAAYKAPALAMELAEVYTGNEP NGTINIEGSGTHPEALOKLHVRADIGTALDGDARVYIVDERGELIVDGLQMAVI ADSVAADNLRGGCIYATVNSNIGLRFGLDGGKLTARTKVDYRVYEHMKNHNVG GEOSGHVLSIDYGTGDGLVALQVLAQVAKRGLIVSEVCKRFEVPVOLLKVRISGG KPLENPVIOAIDAASALANNRRLVIRPSGTPEPLIRVMAEGDDSAKKEKIVNDLVG ISSRSRA"
	complement(1654. .3600)
	/gene="ftsH"
	/note="Atu3710"
	complement(1654. .3600)
	/gene="ftsH"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
	/codon_start=1
	/transl_table=11
	/product="metallopeptidase"
	/protein_id="A144520.1"
	/db_xref="GI:17742132"
	/translacion="MNPFRNALMAVIALMLALPFSMFQTSPTQSGNELPYSGFIR DVSQGRVDTVGNRVLGYTENGTAFCQYSPVIDSILMERLOSKNVIVARPSDGS SGEPLSTGLTLPMLITIGWALFFMRMOGGSGGAPKSAKLTLEHSGVTEDDV ACYDAEADIDELIYELPRDQKRFQKRGKIPROVILLVGPGRGKTLTASVAGNEAV FFTISGSDVEEMFVGASRVKDFEADKANACITITFDEIDAVGHRGAGIGGND ERIOLTLNLMVEMGFEEANEGILIIATNRPDVALPLRGRDROVVAVDTVIGRE RILTVHVRNVLAPNDVILIGRTGFGSCADLMNLVNEPALAARKNRVVDTQIEF DAKKIMGAERSSAMTAEKRLITAYHEGHAITAKVAADPLHKATIIIPRGALG VMVOLLPGDRSMTYKMWYSRLVIIMGGRVAEELTGKKNITSGASDIEQATKLARA MVTQWGSADLQGAENGQEVFLGHVSQSNVSEATRAOITIDEFVRLIDEATEA RRIITLDHNDGVAITABGLTLEYLTGTETIKALLRGKRPARDLGDSDPSGRSGSAVAKG AKKQGPSEAGDGEAGGMEPPH"
	complement(1654. .3600)
	/gene="ftsH"
	/note="Atu3710"
	complement(1654. .3600)
	/gene="ftsH"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
	/codon_start=1
	/transl_table=11
	/product="metallopeptidase"
	/protein_id="A144520.1"
	/db_xref="GI:17742132"
	/translacion="MNPFRNALMAVIALMLALPFSMFQTSPTQSGNELPYSGFIR DVSQGRVDTVGNRVLGYTENGTAFCQYSPVIDSILMERLOSKNVIVARPSDGS SGEPLSTGLTLPMLITIGWALFFMRMOGGSGGAPKSAKLTLEHSGVTEDDV ACYDAEADIDELIYELPRDQKRFQKRGKIPROVILLVGPGRGKTLTASVAGNEAV FFTISGSDVEEMFVGASRVKDFEADKANACITITFDEIDAVGHRGAGIGGND ERIOLTLNLMVEMGFEEANEGILIIATNRPDVALPLRGRDROVVAVDTVIGRE RILTVHVRNVLAPNDVILIGRTGFGSCADLMNLVNEPALAARKNRVVDTQIEF DAKKIMGAERSSAMTAEKRLITAYHEGHAITAKVAADPLHKATIIIPRGALG VMVOLLPGDRSMTYKMWYSRLVIIMGGRVAEELTGKKNITSGASDIEQATKLARA MVTQWGSADLQGAENGQEVFLGHVSQSNVSEATRAOITIDEFVRLIDEATEA RRIITLDHNDGVAITABGLTLEYLTGTETIKALLRGKRPARDLGDSDPSGRSGSAVAKG AKKQGPSEAGDGEAGGMEPPH"
	complement(1654. .3600)
	/gene="ftsH"
	/note="Atu3710"
	complement(1654. .3600)
	/gene="ftsH"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
	/codon_start=1
	/transl_table=11
	/product="metallopeptidase"
	/protein_id="A144520.1"
	/db_xref="GI:17742132"
	/translacion="MNPFRNALMAVIALMLALPFSMFQTSPTQSGNELPYSGFIR DVSQGRVDTVGNRVLGYTENGTAFCQYSPVIDSILMERLOSKNVIVARPSDGS SGEPLSTGLTLPMLITIGWALFFMRMOGGSGGAPKSAKLTLEHSGVTEDDV ACYDAEADIDELIYELPRDQKRFQKRGKIPROVILLVGPGRGKTLTASVAGNEAV FFTISGSDVEEMFVGASRVKDFEADKANACITITFDEIDAVGHRGAGIGGND ERIOLTLNLMVEMGFEEANEGILIIATNRPDVALPLRGRDROVVAVDTVIGRE RILTVHVRNVLAPNDVILIGRTGFGSCADLMNLVNEPALAARKNRVVDTQIEF DAKKIMGAERSSAMTAEKRLITAYHEGHAITAKVAADPLHKATIIIPRGALG VMVOLLPGDRSMTYKMWYSRLVIIMGGRVAEELTGKKNITSGASDIEQATKLARA MVTQWGSADLQGAENGQEVFLGHVSQSNVSEATRAOITIDEFVRLIDEATEA RRIITLDHNDGVAITABGLTLEYLTGTETIKALLRGKRPARDLGDSDPSGRSGSAVAKG AKKQGPSEAGDGEAGGMEPPH"
	complement(1654. .3600)
	/gene="ftsH"
	/note="Atu3710"
	complement(1654. .3600)
	/gene="ftsH"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
	/codon_start=1
	/transl_table=11
	/product="metallopeptidase"
	/protein_id="A144520.1"
	/db_xref="GI:17742132"
	/translacion="MNPFRNALMAVIALMLALPFSMFQTSPTQSGNELPYSGFIR DVSQGRVDTVGNRVLGYTENGTAFCQYSPVIDSILMERLOSKNVIVARPSDGS SGEPLSTGLTLPMLITIGWALFFMRMOGGSGGAPKSAKLTLEHSGVTEDDV ACYDAEADIDELIYELPRDQKRFQKRGKIPROVILLVGPGRGKTLTASVAGNEAV FFTISGSDVEEMFVGASRVKDFEADKANACITITFDEIDAVGHRGAGIGGND ERIOLTLNLMVEMGFEEANEGILIIATNRPDVALPLRGRDROVVAVDTVIGRE RILTVHVRNVLAPNDVILIGRTGFGSCADLMNLVNEPALAARKNRVVDTQIEF DAKKIMGAERSSAMTAEKRLITAYHEGHAITAKVAADPLHKATIIIPRGALG VMVOLLPGDRSMTYKMWYSRLVIIMGGRVAEELTGKKNITSGASDIEQATKLARA MVTQWGSADLQGAENGQEVFLGHVSQSNVSEATRAOITIDEFVRLIDEATEA RRIITLDHNDGVAITABGLTLEYLTGTETIKALLRGKRPARDLGDSDPSGRSGSAVAKG AKKQGPSEAGDGEAGGMEPPH"
	complement(1654. .3600)
	/gene="ftsH"
	/note="Atu3710"
	complement(1654. .3600)
	/gene="ftsH"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
	/codon_start=1
	/transl_table=11
	/product="metallopeptidase"
	/protein_id="A144520.1"
	/db_xref="GI:17742132"
	/translacion="MNPFRNALMAVIALMLALPFSMFQTSPTQSGNELPYSGFIR DVSQGRVDTVGNRVLGYTENGTAFCQYSPVIDSILMERLOSKNVIVARPSDGS SGEPLSTGLTLPMLITIGWALFFMRMOGGSGGAPKSAKLTLEHSGVTEDDV ACYDAEADIDELIYELPRDQKRFQKRGKIPROVILLVGPGRGKTLTASVAGNEAV FFTISGSDVEEMFVGASRVKDFEADKANACITITFDEIDAVGHRGAGIGGND ERIOLTLNLMVEMGFEEANEGILIIATNRPDVALPLRGRDROVVAVDTVIGRE RILTVHVRN

gene
CDS

gene
CDS

gene
CDS

```
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"  
/strain="C58"  
/db_xref="taxon:181661"  
136..669  
/gene="AGR_L_2246"  
136..669  
/gene="AGR_L_2246"  
/note="ompL6 protein - Brucella abortus"  
/codon_start=1  
/transl_table=11  
/product="AGR_L_2246p"  
/protein_id="AKR89692.1"  
/db_xref="GI:15159600"  
/translation="MSRTNTSALSPWOKLARNPAYMTLALAAGCAKKNNPNSAG  
EGLGASGATPGSGODFTYVNGVDRIFFPDITSIRADAQOTIQRPAOWISRYPNAI  
TYEGHDADEGTREYNIALGARRAATRDLFASGVAPSRMKITSYGEKRPVAACDDIS  
CMSQNRRAVTVLGGAQM"  
828..1832  
/gene="AGR_L_2247g1"  
828..1832  
/gene="AGR_L_2247g1"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=11  
/product="AGR_L_2247g1p"  
/protein_id="AKR89693.1"  
/db_xref="GI:15159601"  
/translation="MRLIVAGMLIACTGSITAVSGPLFGAGSGFGTINSQPPI  
VRQAWEAAYVOLEOIEINOLNRIEMSFOLIOMOEITRKAOEDNEFRFOLEGKG  
ASTPTAAKKAASPVNPAPTLPDIDVATIIEPPOGASVPSSSAATAAGETLL  
GFELDSKGPVGVTILNGANNSSGLPEVTINGNRKTDPNVTAITSEGDIIYAAY  
GHVLSDPYLKLEGGFOOYLGYPKCKTKADASFVLGEADYSOGKFNEAKTFUNGQT  
YGRSKRAPAMIKIGNSTLAALNDTEFACTLRLEVPKRYPNASKTYLSKVASEOKRLSC  
1857..3188  
/gene="AGR_L_2252"  
1857..3188  
/gene="AGR_L_2252"  
/note="(AJ243808) Yaen protein"  
/codon_start=1  
/transl_table=11
```


/gene="dnak"
 /note="CT0643"
 complement(4720..6621)
 /gene="dnak"
 /note="identified by match to PFAM protein family HMM PF00012"
 /codon_start=1
 /transl_table=11
 /product="Dnak protein"
 /protein_id="AAW71882.1"
 /db_xref="GI:21646580"
 /translation="MGKIGIDIGTNSCAVVMGCTPTVIENSSEGRPTPSMAVFTK
 TGERLVGOAKRQAVNPKRTITSIKREGRKRYDEPNKRTIASDYVNEGAYAVKI
 GDRYSPQETSAITLQMKOTADFLGEKYTEVITVPYFNDAGQATKDKAGTAGL
 EVKRIINEPTAAALAYGLDKRKENERVAADFGLGFFDISLEGGVEVSTDDI
 HEGDDPDDVITVYLADEPKKQEGIDLRDAIALQRLKRAAEKAKIELSSRTDTEINL
 PFTAQOEGKRLVINTLRAKFEAMSAALFDLKEPCRAIKSKFDIEIDIEVYLVG
 GSTRIPVQALVKEFFGKEPNKSNPEVVAIGALOGGVLOGDVTDLVLTPLSL
 GFTLGVMKLTLEANTTPTPKQEIESTAADNQTVEVHVLQGEPRMADSKTIGRE
 HSDTPPAPRGVQIETPTDIDANGILNLSAKRATGKQSGIKIESGKTAELEIK
 KEDAKAAAEQKREBELEKNSADSLITSEKQTELGDKLPADKKAIESALEKIK
 EAKSGRVDAIKPAMDELKRVMSDASNLVGPAPPEPPTNGHAGSGSGGGAANA
 YEVIDGDDK"
 complement(6701..7099)
 /gene="CT0644"
 complement(6701..7099)
 /note="identified by match to PFAM protein family HMM PF00011"
 /codon_start=1
 /transl_table=11
 /product="heat shock protein, Hsp20 family"
 /protein_id="AAW71883.1"
 /db_xref="GI:21646581"
 /translation="MLMKIAKDPMLFDDIWSSQMAVAPSEKVIDISEDNAYHDAE
 LPGAKEQIALNIEDVLTITKERTHEEKKNYHVERTYGSFSRSENGEIIQOE
 HIGATVNGVLTHTLPKTPAKTKKEIPIN"
 7118..7228
 /gene="CT0645"
 7118..7228
 /note="identified by glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAW71884.1"
 /db_xref="GI:21646582"
 /translation="MFVFSRFGMLRSILTSVITSKDVPKRFHFDLCV"
 complement(7467..7856)
 /gene="CT0646"
 complement(7467..7856)
 /gene="CT0646"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator, ArsR family"
 /protein_id="AAW71885.1"
 /db_xref="GI:21646583"
 /translation="MSKSMRTTEEVKKKQFNMPEMDLDAVSNFKLLSPMKILR
 ALCHREHTVQIVEGASQANISKHIALMHNDNVNRRKGLKCYIRISDDSVYAC
 FLISKSVENLQDRISLWQKYNLTNT"
 complement(8023..8250)

Alignment Scores:

Pred. No.: 5,82e-17 Length: 10643
 Score: 259.50 Matches: 56
 Percent Similarity: 56.41% Conservative: 32
 Best Local Similarity: 35.90% Mismatches: 63
 Query Match: 30.14% Indels: 5
 DB: 1 Gaps: 2

US-09-674-779b-2 (1-172) x AE012837 (1-10643)

QY 15 SerValLeuThrPheMetThrGlyCysAlaIaanslySerThrSerGlnValMetValAla 34

Db 1864 GCGGCATGATCATTTTGGCAGCGCTGTCTCAAAAAGTGA-----GTATCAACAGAT 1917
 QY 35 ProAnaIaProthmGlyTyrThyGlyValIleYrThrGlyVal-----AlaPro 51
 Db 1918 GAACCTTCAGAGCGCGTTATGATCGGATGGAGCGGCACTGAGACAGCAGCAGA 1977
 QY 52 LeuValAspAnaAspGluThrValIleAlaLeuAlaLeuAseryLeuProSerLeuValTyr 71
 Db 1978 GTGACGTGGAGAGATATCGCCCAAGGCGCAAGCCCGTTCATATTTGGATATTTTC 2037
 QY 72 PheAspPheAspSerAspGluIleLysFroGlnAlaAlaAlaIleLeuAspGluAla 91
 Db 2038 TTCGATTTCGACAGTTCGCGCACTAAGCTCCAGACAGCAAGAACATCAACAAAGCC 2097
 QY 92 GlnPheLeuThrThaAsnGlnThrAlaArgValLeuValAlaGlyVHISThrAspGluArg 111
 Db 2098 GCCTGGATGCAGAAAAACCTACTCTGCGGTATCATTTAGGAGCATTGGCATGACGCG 2157
 QY 112 GlySerArgGluTyrAsnMetSerLeuGlyLuuArgAlaValAlaValArgAsnTyr 131
 Db 2138 GGCACCGATGACTACACATCGCGCTGTAGCGTGTCTGAAGCAGCAGATGTAT 2217
 QY 132 LeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArg 151
 Db 2218 CTGTTTAACTTGAGTAAAGCGAGCGCGCTTCTTCTTACCGTCAGTTTACGAGGAGAAAG 2277
 QY 152 ProIleAlaPheGlyThrAsnGlnGluIleArgSerGlnAsnArg 167
 Db 2278 CCGTTTGACCCCGGTCATACAGAGGCTTGGGCAAGAACAGAGA 2325
 RESULT 37
 AE004259 11914 bp DNA linear BCT 31-JUL-2000
 LOCUS Vibrio cholerae chromosome I, section 167 of 251 of the complete
 DEFINITION
 accession AE004259 AE003852
 version AE004259.1 GI:9656353
 keywords
 SOURCE
 ORGANISM
 Vibrio cholerae.
 Bacteria; Proteobacteria; gamma subdivision: Vibrionaceae; Vibrio.
 REFERENCE
 AUTHORS
 1 (bases 1 to 11914)
 Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
 Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
 Gill, S.R., Nelson, K.E., Read, T.D., Tetelin, H., Richardson, D.,
 Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I.,
 Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D.,
 Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
 Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
 Direct Submision
 Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1..11914
 /organism="Vibrio cholerae"
 /strain="N16961"
 /serotype="O1"
 /db_xref="taxon:666"
 /chromosome="I"
 /note="biotype: El Tor"

gene
89. .1255
/gene="VC1827"
CDS
89. .1255
/gene="VC1827"
/note="similar to GB:M15380 SP:P00946 PID:146722 GB:000096
PID:1742663; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="mannose-6-phosphate isomerase"
/protein_id="AA094975.1"
/db_xref="GI:9656354"
/translation="MTNPIONYANGSTALQQLFELENPDDEQALIMGAPHGSCG
IANGKIVQSDFTAQNPNLIGETITARQGELEPYEKILAAEKALSIQHPKQOAE
RGEFAQESQGLALTAQORNYKDPNHKPELVYALTEYQAMGFRANOEILNYFIELSD
EIOPLVWQSNPTQGLRDFEGLSLQGEAKRBALEALMAQKOIDLELPOLIVEL
ENQYNDIGLEFAPLMNVITLQGEAMFLDAETPHATLHGTGLIEMANSNVNACIT
PKYMDIDELVACQGFKKHPLKLEOLRKEFEVGEQGVIPADKRALIIPPTQOTIGV
KSAFVPLDSMSLRHANGEOCLVRKQGVIFPAIAEOYTIICNGKVARAFSA"
complement(1375. .1827)
/gene="VC1828"
CDS
complement(1375. .1827)
/gene="VC1828"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA094976.1"
/db_xref="GI:9656355"
/translation="MKARGPCIRHAAADFAFVWFCFVGMIIELFVSGMTPQOGLA
SRILSVINIAIAMPYGVPRDYLROGRKISPTGMMKNLSDLVAYIFQSPVYATILE
TVGASTDQITTAATNALVSCGMGVLTGYFLDMCRMFKPGYVSEG"
complement(2016. .2123)
/gene="VC1829"
CDS
complement(2016. .2123)
/gene="VC1829"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA094977.1"
/db_xref="GI:9656356"
/translation="MWPSSAGKLSHRAHADVTKLPNVNLETDVLTAKNA"
2101. .2196
/gene="VC1830"
CDS
2101. .2196
/gene="VC1830"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA094978.1"
/db_xref="GI:9656357"
/translation="MPALGHTHKKGKKEGRNRVIYSTDGKDSR"
complement(2163. .4373)
/gene="VC1831"
CDS
complement(2163. .4373)
/gene="VC1831"
/note="similar to GB:007069 PID:459205 SP:P54302;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="sensor histidine kinase"
/protein_id="AA094979.1"
/db_xref="GI:9656358"
/translation="WMAIRSSILKKRSIIALTITLAFPLAIVGLTSLWGLVEPPEKEL
KNMLRAELATQIREPLNLSIGVLOSITSGSADKEBERMLRSLSFVYGVIT
SGLMPEPMLSATDPSLRYSLSFPNKATDGOVDOSNNRPRAGCYDREDSWTLAERE
AGIYVSPVYDPTRYPMITVSTPIYNGCFAGVATVDSLESLLQFVATAEOYN
LGNLKDAGVGVSHNRTYNALVSYISSEFEWMQLEVVANQHVDEIIFDLIINI
EGWMLILLCVAGVFILSHYMRPIYLAKEVSEREGELIDIRKSOEIRHLID
TFNQKITYLEAKVRAQASTKAKSAFLATLSHEIRTPMNGVGTAOILLKDELTKOR
OHLTSLYSEGEHMSLNLNLIIDYSKIEOGKFLDSAPLKSIIIGISIKSYSSICVEK
GLKFOINSEITDGRWYGDKARLRQIIFNLISNAVKFIENGAVAGLSDESCDENYI

gene
I1KVQDTGIGIAQESIGRIFRPEQAEHTTTRRGCTGICAIYKQIAELANGTVLVQ
SEVQGSCKRVKRLATIEPVEDVKPIAKTYPGILRVILVEDENRFTMLLEAFMRNK
GEECHSVMDGVQVATLALQESSFDLIMQNDKKGIGIATREIRQLPPOAKILLFGC
TADVEFKDTRDKMISAGADDIIAKPIAEHELDMLRQHSRLTYQFHRSPSLVSE"
4581. .5153
/gene="VC1832"
CDS
4581. .5153
/gene="VC1832"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA094980.1"
/db_xref="GI:9656359"
/translation="MOTLOELIQPELENKLLNQAQTVGFYVTAACCPNVLPEEWLP
FLMGGEIAPFSDPIOLENVEFHHIALNMTESQLENNMOPFTGLDEAEIYQOEY
RDFCBMQLQMLADMDWELVLPENSSNDLGLVGLSLMLYDEPSTIATLYEGGFT
GLEQFAEIVHAIIPVMLCGITRGVTLAEQ"
complement(5227. .6288)
/gene="VC1833"
CDS
complement(5227. .6288)
/gene="VC1833"
/note="similar to SP:P1458 GB:X12713 PID:581137 GB:000096
PID:1651334; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="guinolate synthetase A"
/protein_id="AA094981.1"
/db_xref="GI:9656360"
/translation="MSHLIDTINLYPPPPKPIPLRDEKOAYIAETIKOLLERDAVL
IAHYTDEPIQALAESTGEFVDSLEMAKFGGRYPATYLLIAGRFMESAKITLPEK
RLIMPLEIEKESLIDGCPADKPERECDDPHRRVYANTNSAKAKRADRVYTSITLAL
EIVEHIDSEGRPIINGPRRHGATYAKTGADMLMOECVVAHDEFSDALRKXKALY
PDAILVLPESPASVVELADAVGTSOLIKAKTLPQOKMIVADDKGIFFKMOQWVE
KELIAPATAGAGATSCAHCAPWAMMNGLOAIAQALREGGQHEHIFVEALRVSLIP
INRMLDFAEOLNLKYGKNA"
complement(6462. .7226)
/gene="VC1834"
CDS
complement(6462. .7226)
/gene="VC1834"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA094982.1"
/db_xref="GI:9656361"
/translation="WFSNLKHVVMMLTASASYALAAPVSDLGSNVSGSRSSEI
ERLERLESRLNIVLOMQKOLDMALFVNELRQIENKNYDMQOMLROBELFVLEDR
VRGEKSPSSGSGOLPTSSNDEAAGFSSANQOAYQONAVDMLIKRKYAGAIAR
OKFOFDYNSFTFSANAHYMLGOLYFAKKEDEKAKSFAAAYSDKSNKRADALVKLD
IAKRNNNAEOKARKEVQCAVDPEYPPSASAKIAKENIK"
complement(7242. .7760)
/gene="VC1835"
CDS
complement(7242. .7760)
/gene="VC1835"
/note="similar to SP:P07176 PID:41360 PID:42257 GB:000096
PID:1651333; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="peptidoglycan-associated lipoprotein"
/protein_id="AA094983.1"
/db_xref="GI:9656362"
/translation="MOLNKYLSLIIALPIVATYACSSSDAANSQSQTQSAVSYTD
TIECHADBERGPEPNIHALGERRQAVAKYLEALGVQGLSYSGEKLVLVGQSEB
AYANRAAVLVY"
complement(7792. .9144)
/gene="VC1836"
CDS
complement(7792. .9144)
/gene="VC1836"
/note="similar to GB:M28232 SP:P19935 PID:148020 GB:000096
PID:1651332; identified by sequence similarity; putative"
/codon_start=1

DB	442	TTAGCATAC	450
RESULT 40			
AE006136/c			
LOCUS	AE006136	11061 bp	DNA linear BCT 08-MAR-2000
DEFINITION	Pasteurella multocida PM70 section 103 of 204 of the complete genome.		
ACCESSION	AE006136	AE004439	
VERSION	AE006136.1	GI:12721286	
KEYWORDS			
SOURCE			
ORGANISM	Pasteurella multocida. Pasteurella multocida bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Pasteurella.		
REFERENCE	1 (bases 1 to 11061)		
AUTHORS	May,B.J., Zhang,Q., Li,L.T., Paustian,M.L., Whittam,T.S. and Kapur,V.		
TITLE	Complete genomic sequence of Pasteurella multocida, PM70		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)		
MEDLINE	21145866		
PUBMED	11248100		
REFERENCE	2 (bases 1 to 11061)		
AUTHORS	Zhang,Q. and Kapur,V.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA		
FEATURES			
Source	Location/Qualifiers		
	1..11061		
	/organism="Pasteurella multocida"		
	/strain="PM70"		
	/db_xref="taxon:747"		
gene	complement(67..357)		
	/gene="PM0962"		
CDS	complement(67..357)		
	/gene="PM0962"		
	/codon_start=1		
	/transl_table=1		
	/product="unknown"		
	/protein_id="AAK03046.1"		
	/db_xref="GI:12721287"		
	/translation="MIFYYDKMSLPVLLMALMLFSGIFLSTQOVLXYORARIYIQ HVQAQLAQNRKORQFGLACENNVRNDLFOICTSQHKVSYSLAELSL"		
	/complement(341..988)		
	/gene="PM0963"		
	complement(341..988)		
	/gene="PM0963"		
	/codon_start=1		
	/transl_table=1		
	/product="unknown"		
	/protein_id="AAK03047.1"		
	/db_xref="GI:12721288"		
	/translation="MAYHKGMVFLVILLSSGLFIIMLNDLILHHLSALHQRKID VQQLVQGRNEKKQHLCKHSSTSSEKROQVFRILEGKANTVWCICRLTKRQPR QALNIGLALFIOREYRGLESHQFSVQAHVPFYVDNFKQTSMTLTONIEAIITAEGLD HIEHKGIKGTIVTGGRLTLEPSOVSYHKATVAWYQLRDMQLAEKSWHDFLL"		
	/complement(985..1698)		
	/gene="PM0964"		
	complement(985..1698)		
	/gene="PM0964"		
	/codon_start=1		
	/transl_table=1		
	/product="unknown"		
	/protein_id="AAK03048.1"		
	/db_xref="GI:12721289"		
	/translation="MLFRGYTLSLMISLSILLVATAFVYVDYRONRAHFMQ LQALELVNIOGMKQDIRAGFRAPETPNKNYPSLPSSEFQASLQPSIHGWNVSCI LRFYEDLMQVCSGGKREKTCIKDGRNNTSHUERLDEYIRLNNGMETRLTYSAVNS RCLALERTYLOOSACHSGGWDLDDDEYIISDLSPFWLMGPALEYILKGYLKKOP HFQYETTWAPLLNHEKQ"		
	/complement(1702..2229)		
	/gene="PM0965"		
	complement(1702..2229)		
	/gene="PM0965"		
CDS			

```

/ gene="PM0965"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAK03049.1"
/ db_xref="GI:12721290"
/ translation="MKRGFTLEILMLVLLISMLLVLPNMTRVIDFISPEOROL
WIFRLQJTRVATSCGVWFLINRPNYNNQKHLCTQKSEYICDCEFAPOHCPORLSNO
FYSSHTGYTMKTHYIPNTRLSAVRDLETTCFVLQDYQKATISLNVGSIKL
KKEASLSACEHNKEN"
/ complement(2910. .3362)
/ gene="PM0966"
/ complement(2910. .3362)
/ gene="PM0966"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAK03050.1"
/ db_xref="GI:12721291"
/ translation="MKRLTKVLLVGSVAVLAAGSSKKDESAGOMFGYSVODLQOR
YNTVYGFPRKYNIEGTYOILDHAHAFNLATPATKYVYEGNTDERGTPYNTALQOR
ADAVKHVLSAKGVQAGVSTVSIGEEKPAVLGHDEAVSKNRRAVLAV"
/ complement(3383. .4681)
/ gene="PM0967"
/ note="PM0967"
/ complement(3383. .4681)
/ gene="PM0967"
/ codon_start=1
/ transl_table=11
/ product="TolB"
/ protein_id="AAK03051.1"
/ db_xref="GI:12721292"
/ translation="MKLYRMWSILIVFLLAVLOPAQAEVRIVIDEGVDSAPRIAVP
FKMNGSAPADTADITLADLRSGKFNPIPVNRMQHTSYSEVPEAMALGIDAV
VGHVTPANGSFNIAVLQVLTIGATGTAGVLTQNTYTAKLRTGATTVDEPEK
LTGIRGAFRTIRIAYIVQKHSQPYEVRSDVDGFPQVNNESAOLPMSHOGK
LAYVSENNKSQLVODLGSARKVVAASFNGHGAFAFSDGSRLLAFASKGVNLTY
VMNLSGOISLTSGTGNNTPEPSPDGQITVFTSDRSGSPQVYOMASAGGASLST
GGRSYSGMADGSAITIMISRONIVKRLDASGSTVLESTFIDENPSISPNCIMITYS
STQGLKVLQDVASDGRFKARLPDNGGVKPEPMSVYLTK"
/ complement(4755. .5924)
/ gene="TolA"
/ note="PM0968"
/ complement(4755. .5924)
/ gene="TolA"
/ codon_start=1
/ transl_table=11
/ product="TolA"
/ protein_id="AAK03052.1"
/ db_xref="GI:12721293"
/ translation="MOSKONNKDISAVILSVLLHALLFGLILIGSLYHRYELMGSGEG
DSEYGAIVVDYDAAQOEMGRLOOQKGGTDKAKKTEPVIEEKVPQDEQEIIRQEI
EKQKALEKQKIEIOKKEIEROKOLAEOKROEORARLALEKQKQAEAKAKOLAEAK
LKAEAEKKRLAALAKQKQEEAKAKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKK
AEKAKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKK
TRGSNTDROGSGGALGAGDGKGTDOYAGVIKRIORFLKEPSFAKVCVEE
FLRDGTIANRYRVSPPDICOAAVSAVARTKVPAPVTDIDYORYKKSPIEFKLR"
/ complement(5940. .6362)
/ gene="TolR"
/ note="PM0969"
/ complement(5940. .6362)
/ gene="TolR"
/ codon_start=1
/ transl_table=11
/ product="TolR"
/ protein_id="AAK03053.1"
/ db_xref="GI:12721294"
/ translation="MSYRKRKRDYKSEINIVPFLDVLVLLLFPMATPITISQVOVE
LPDAVESQSVNEDKRVVILEVAGIGQYSLAIGTERLEGTEEMVYOLSKQEDKDN
TLPLVAGARQVDEVIKALNLLHLAGIKSVGLMTNPI"
/ complement(6438. .7127)
/ gene="TolO"
/ note="PM0970"

```

```

CDS      complement(6438. .7127)
/ gene="TolO"
/ codon_start=1
/ transl_table=11
/ product="TolO"
/ protein_id="AAK03054.1"
/ db_xref="GI:12721295"
/ translation="MTADLNLELFLFKASIVQVFIIVLIIIFSIMSAIIIOISRLIT
KALKDAAFERFWSGEDLNRLFDGLENNKRDGLSGEQLFYVGFKEFSRLKQANPNAP
EATIGSSRAMNLANREVELEFRVPLATVASISPYIGLFTGVGIMHAFALASGA
KQATLQVAVAGTAEALATATIGLFAAIPAVMAVNRSLRSKLEQYKGNIDFETTL
HROAFKNGTH"
/ complement(7146. .7556)
/ gene="PM0971"
/ complement(7146. .7556)
/ gene="PM0971"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAK03055.1"
/ db_xref="GI:12721296"
/ translation="MVANVFNEPIRVYVEDTDAGVYVYHARYLHFERARTEYLRCAG
FSOHTLILNEQLAPVYKTMINDRYPAKLNDLIVETLVTELKAGASIFFQRLKKEO
ILCBATYVACVYDGNKPKPIAIPAKAVLVNEL"
/ complement(7686. .7973)
/ gene="PM0972"
/ complement(7686. .7973)
/ gene="PM0972"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAK03056.1"
/ db_xref="GI:12721297"
/ translation="MINSLYOIFNKGSLRTLSFIAILLTLEFFENINDFSTNRSAP
APLILMTGGLIMHIGEIFEISITLWKGLFSPIDYIGFIIAIVSMAG"
/ complement(8089. .9225)
/ gene="CydB"
/ note="PM0973"
/ complement(8089. .9225)
/ gene="CydB"
/ codon_start=1
/ transl_table=11
/ product="CydB"
/ protein_id="AAK03057.1"
/ db_xref="GI:12721298"
/ translation="MIDYELRFVWMLIIIVLLIGFAYTDFPMGCVTLPLPFAKKEY
EKRIIMINSIAPHMDGNQVWLTAGGAMPAAPIYAVISGFTYIAMILLVLALEFVY
GLEVYRAKIDNPRTMLMDGLFIFGFPVSLVGYAFGLMLOGVPEENELVQVYTS
FPAILNPPALIGIVSLAMLITOGSANTLOMTSELRHARSITQAGAVATIIIFVLA
GVWLYFRNGPVTYSVIDHNASSTPRGKVAIEAGANNYKMDPLMIFPLAVSAV
LNLILASKDRSGFALESLSIMAGIILLTAGISMPFVMSITHPLESLIMMDSISKL
TLLTMLGFAVFEVVLVLLAVTTIWAIVKMFGRIDSOPIEHNKSLY"
/ complement(9240. .10802)
/ gene="Cyda"
/ note="PM0974"
/ complement(9240. .10802)
/ gene="Cyda"
/ codon_start=1

Alignment Scores:
Pred. No.:      8, 82e-17      Length:      11061
Score:          258.00      Matches:      50
Percent Similarity: 66.998      Conservative: 19
Best Local Similarity: 48.548      Mismatches: 34
Query Match:    29.97%      Indels:      0
DB:             1      Gaps:      0

US-09-674-779B-2 (1-172) x AB006136 (1-11061)
QY      70 ValTYPhaSpPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGlu 89
Db      3221 GTGTATTTCGCGCTTCGATAAATACATATCGAAGGTGAATATGTAACAATTTTATGATCA 3162

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:16:20 ; Search time 215 Seconds

(without alignments)
1801.599 Million cell updates/sec

Title: US-09-674-779B-2
Perfect score: 861
Sequence: 1 MMLHIQIAAAAAALSVLTFFM.....IAFGTNEBAMSONRRRELSTY 172

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Command line parameters:
-MODEL=frame+ .p2n.model -DEV=xlp
-O=/cqr2.1/USPTO/spool/US09674779/runat.06072003_121613.9795/app_query.fasta_1.327
-Db=N.Geneseq.101002 -QFMT=fastlap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi
-LIST=100 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09674779.@CGN_1.1.208.@runat.06072003_121613.9795 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

N.Geneseq.101002.*
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	861	100.0	519	21 AAZ40351	M. catarrhalis BAS
2	861	100.0	92407	22 AAF28549	Genomic fragment #
3	855	99.3	519	21 AAZ40352	M. catarrhalis BAS
4	850	98.7	519	21 AAZ40353	M. catarrhalis BAS
5	844	98.0	519	21 AAZ40354	M. catarrhalis BAS
6	269.5	31.3	866	11 AAQ06089	Sequence encoding
7	264	30.7	737	11 AAQ03869	H. influenzae DNA f
8	264	30.7	867	9 AAN81194	16600 dalton outer
9	264	30.7	1019	15 AAQ45440	Fragment encoding
10	264	30.7	1830121	17 AAT42063	Haemophilus influe
11	259	30.1	737	9 AAN80226	Sequence of Haemop
12	215.5	25.0	540	18 AAT67789	H. pylori outer me
13	215.5	25.0	540	18 AAT77469	H. pylori outer me
14	215.5	25.0	540	22 AAF25593	H. pylori HPS14 e
15	215.5	25.0	561	18 AAT68048	H. pylori outer me
16	211.5	24.6	770	20 AAV90653	Nucleotide sequenc
17	211.5	24.6	1236	20 AAV90843	Nucleotide sequenc
18	211.5	24.6	1610	20 AAV90545	Nucleotide sequenc
19	205.5	23.9	435	18 AAT74194	H. pylori Omp22 ge
20	204.5	23.8	579	24 ABL91202	Chlamydia pneumoni
21	204.5	23.8	779	22 AAD08593	Chlamydia pneumoni
22	204.5	23.8	1230025	20 AAX91990	Nucleotide sequenc
23	196.5	22.8	1038602	20 AAX01425	Complete genome se
24	180.5	21.0	675	22 AAF30043	Moraxella catarrha
25	180.5	21.0	31940	22 AAF28526	Genomic fragment #
26	176.5	20.5	1830121	17 AAT42063	Haemophilus influe
27	175.5	20.4	645	17 AAT32600	P. aeruginosa OprF
28	172	20.4	4274	14 AAO54144	Sequence of Plasmid
29	171.5	19.9	486	17 AAT32599	P. aeruginosa OprF
30	171.5	19.9	681	17 AAT34419	P. aeruginosa OprI
31	171.5	19.9	1253	9 AAN82023	Outer membrane pro
32	170	19.6	1319	21 AAC38558	Actinobacillus ple
33	169	19.6	1110	21 AAC79664	Actinobacillus ple
34	165.5	19.2	1720	16 AAO78916	Non-typable Haemop
35	165	19.2	898	24 AAT46503	M. catarrhalis MCA1
36	165	19.2	96109	22 AAF28548	Genomic fragment #
37	161.5	18.8	1035	24 ABA91419	Haemophilus paraga
38	161.5	18.8	1035	24 ABA91422	Haemophilus paraga
39	161.5	18.8	1035	24 ABA91423	Haemophilus paraga
40	161.5	18.8	1035	24 ABA91425	Haemophilus paraga
41	161	18.7	1026	24 ABQ90179	Haemophilus paraga
42	160	18.6	981	22 AAF23499	M. capsulatus gene
43	158.5	18.4	1035	24 ABA91417	C. coli Cadr DNA.
44	158.5	18.4	1035	24 ABA91420	Haemophilus paraga
45	158.5	18.4	1035	24 ABA91420	Haemophilus paraga
46	158.5	18.4	1035	24 ABA91426	Haemophilus paraga
47	155.5	18.1	1026	24 ABA91421	Haemophilus paraga
48	155.5	18.1	1026	24 ABA91424	Haemophilus paraga
49	153.5	17.2	1026	24 ABA91427	Haemophilus paraga
50	148.5	17.2	782	21 ABA81540	N. meningitidis pa
51	148.5	17.2	6477	11 AAQ02030	Plasmid pluf 4-49.
52	148.5	17.2	349980	21 AAF21607	Neisseria meningi
53	148.5	17.2	1437668	21 AAN81490	N. meningitidis B
54	147	17.2	66788	23 AAS59515	Propionibacterium
55	145	16.8	861	22 AAF23500	C. jejuni cadr-M129
56	145	16.8	861	22 AAF23501	C. jejuni cadr-F380
57	144	16.7	669	20 AAX91705	Porphyromonas ging
58	144	16.7	690	20 AAX91580	Porphyromonas ging
59	143.5	16.7	5579	11 AAQ02032	Plasmid pHS 164.
60	143	16.6	4403765	22 AAT99683	Myobacterium tube
61	143	16.6	4411529	22 AAT99682	Myobacterium tube
62	142.5	16.6	1095	21 AAC79663	Virulence gene #70
63	142	16.5	2016	20 AAX91718	Porphyromonas ging
64	142	16.5	2037	20 AAX91591	Porphyromonas ging
65	138.5	15.9	1922	21 AAX38557	Actinobacillus pie
66	136.5	15.9	411	24 ABA83312	Partial OmpA codin
67	136.5	15.9	691	24 AAT46497	M. catarrhalis MCA1

68	136.5	15.9	730	17	AAT28425	K. pneumoniae dete
69	136.5	15.9	730	22	ABA76835	Klebsiella pneumon
70	136.5	15.9	1008	16	AAT03490	K. pneumoniae p40 g
71	136.5	15.9	1008	17	AAT31607	Klebsiella pneumon
72	136.5	15.9	1008	18	AAV13867	Gene coding for P4
73	136.5	15.9	1032	20	AAZ30477	K. pneumoniae OmpA
74	136.5	15.9	1032	22	AAH78461	Nucleotide sequenc
75	136.5	15.9	1032	22	AAF90077	Nucleotide sequenc
76	136.5	15.9	1032	22	AAF80152	Nucleotide sequenc
77	136.5	15.9	1035	18	AAV13862	Gene coding for LP
78	136.5	15.9	1035	21	AAV13868	DNA encoding a P40
79	136.5	15.9	1035	21	AAV15036	DNA encoding a K1e
80	136.5	15.9	1035	21	AAV15036	CDNA encoding a P4
81	136.5	15.9	1035	21	AAV15036	DNA encoding an ou
82	136.5	15.9	1035	21	AAV15036	DNA encoding an ou
83	136.5	15.9	1035	21	AAV15036	DNA encoding a K1e
84	136.5	15.9	1035	22	AAH74731	Nucleotide sequenc
85	136.5	15.9	1035	22	AAH74731	Nucleotide sequenc
86	136.5	15.9	1035	22	AAH74731	Nucleotide sequenc
87	135.5	15.7	640681	24	ABA92787	Branhamella catarr
88	135.5	15.7	640681	24	ABA92787	Buchnera sp. genom
89	130.5	15.2	2251	22	AAH85140	R. anatisperifer o
90	129	15.0	269223	22	AAH85140	Genomic fragment #
91	122.5	14.2	1650	22	ABA89117	M. capsulatus gene
92	122.5	14.2	28989	22	ABA89112	Escherichia coli p
93	117	13.6	1740	23	ABL53296	2-ketoadonate red
94	117	13.6	1788	23	AAH89925	DNA encoding novel
95	117	13.6	1788	23	AAH89925	DNA encoding novel
96	117	13.6	1788	23	AAH89925	DNA encoding novel
97	117	13.6	1788	23	AAH89925	DNA encoding novel
98	117	13.6	2377	23	AAH77370	DNA encoding novel
99	117	13.6	21948	20	AAH20562	Polynucleotide seq
100	115	13.4	242	20	AAV90680	Nucleotide sequenc

ALIGNMENTS

RESULT 1
ID AAZ40351 strand: DNA; 519 BP.

XX AAZ40351;

DT 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein coding sequence #1.

XX BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
XX upper respiratory tract infection; middle ear infection; therapy; ss.

OS Moraxella catarrhalis.

PN WO957277-A2.

PD 11-NOV-1999.

PF 03-MAY-1999; 99WO-EP03038.

PR 06-MAY-1998; 98GB-0009683.

PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Ruelle J;

DR WPI; 2000-062148/05.

DR P-PSDB; AAV55089.

PT Novel BASB019 polynucleotides and polypeptides from Moraxella
catarrhalis used to prepare vaccines against bacterial infections

PS Claim 10; Fig 2; 101p; English.

XX This sequence encodes a Moraxella catarrhalis BASB019 protein of the
CC invention. The sequences can be used for diagnosis of disease, staging of
CC disease, or determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and for
CC screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB019 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies, and as a
CC target for the screening of antimicrobial drugs. The polypeptides can
CC also be used in vaccine formulations, and to identify agonists and
CC antagonists. The polypeptides, antibodies, agonists and antagonists
CC (which are bacteriostatic) are used for the treatment and prevention of
CC diseases including bacterial infection, otitis media in infants and
CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of in-dwelling devices or by other surgical techniques.
CC The frequency of Moraxella catarrhalis infections has risen dramatically,
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.

XX SQ Sequence 519 BP; 139 A; 108 C; 122 G; 150 T; 0 other;

Alignment Scores:

Pred. No.:	7,59e-96	Length:	519
Score:	861.00	Matches:	172
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-674-779b-2 (1-172) x AAZ40351 (1-519)

QY	1	MetMetLeuHisIleGlnIleAlaAlaAlaAlaLeuSerValLeuThrPheMet	20
DB	1	ATGATGTTACATATTCATCAATTCGCGCGCGCTTATCGTATCATCTTTATG	60
QY	21	ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly	40
DB	61	ACAGGCTGTGCAATTAATCAACAGTCAAGTTATGCTCTTAATGACCCACAGCT	120
QY	41	TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys	60
DB	121	TACACGCGGGTTATCTATCTGCTGTGACCTTGTGTGATATGATGACACCGTTAAG	180
QY	61	AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspIleLys	80
DB	181	GCTCTGCAAGCAAGTACCACTGTTGTTTATTTTGACTTGATCTGATAGATTAA	240
QY	81	ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla	100
DB	241	CCGCAAGCTGCTGCCATCTTACGACCAACACAAATTTTAAACCAACATCAACAGCT	300
QY	101	ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu	120
DB	301	CGTGTGTTGGTTCAGATCATACCGATGAGCGGTGATGCGATATATATGTCACAG	360
QY	121	GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla	140
DB	361	GGGGAACGCGCTGCGTGGCGGTCACCACTATTGCTTGTAAGGCAATTAATCAAGCC	420
QY	141	SerValGlnIleIleSerPheGlyGluGluValGProIleAlaPheGlyThrAsnGluGlu	160
DB	421	AGCGTTGAGATTATACGTTTGGTGAAGACGCCCTATGCAATTGGCACAAATGAGAA	480

QY 161 AlaTSPserGlnAsnArgAlaGluLeuSerTyr 172
Db 481 GCATGTCACAAATCGTCTGCTGCAACTGCTTAT 516
RESULT 2
ID AAF28549 standard; DNA: 92407 BP.
XX AAF28549;
AC AAF28549;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #36.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KM bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO200078968-A2.
XX
PD 28-DEC-2000.
XX
PE 16-JUN-2000; 2000WO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
XX
DR WPI; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX
PS Claim 1; Page 369-391; 545pp; English.
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28549-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 92407 BP; 26788 A; 17581 C; 20150 G; 27888 T; 0 other;
Alignment Scores:
Pred. No.: 1.03e-92 Length: 92407
Score: 861.00 Matches: 172
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-09-674-779b-2 (1-172) x AAF28549 (1-92407)
QY 1 MetMetLeuHisIleGlnIleAlaIleAlaIleAlaLeuSerValIleuThrPheMet 20
Db 26057 ATGAGGTATACATATTCATATTCGCCCGCTGCGCGCTTATCGTACTACTTTATG 25998
QY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
Db 25997 ACAGGCTGTGCCAATAATCAACAGTCAAGTTATGTGCTCAATGACACCCACAGT 25938
QY 41 TyrThrGlyValIleIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrVallys 60

Db 25937 TACACTGGGCTTATCTATACATCGTGTGCACCTTTGGTAGATTAAGATGACACCGTTAAG 25878
QY 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80
Db 25877 GCTGTGGCAGACACTACCCAGTTGGTTATTTGACTTTGATTCGATGACATTAA 25818
QY 81 ProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
Db 25817 CCGCAAGCTGCGCCATCTTAGACACACAGACAACTTTTAACCAACCAATCAACAGCT 25758
QY 101 ArgValLeuValAlaIleHisThrAspGluArgGlySerArgGlyIleAsnGlnAla 120
Db 25757 CGTGTGTTGGTTGGCAGGTCATACCATAGAGGTGGTACTCGTAGTAAATATATGCACTG 25698
QY 121 GlyIleuArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
Db 25697 GGGGAACGCCCGTGGCGGTAGCGCAACTATTGCTGTGTAAGGCAATTAAATCAAGCC 25638
QY 141 SerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu 160
Db 25637 AGCGTTGAGATTATCAGTTTGGTGAAGAACGCCCTATCGCATTTGGCACAAATGAAGAA 25578
QY 161 AlaTSPserGlnAsnArgAlaGluLeuSerTyr 172
Db 25577 GCATGTCACAAATCGTCTGCTGCAACTGCTTAT 25542
RESULT 3
ID AAZ40352 standard; DNA: 519 BP.
XX AAZ40352;
AC AAZ40352;
XX
DT 01-MAR-2000 (first entry)
XX
DE M. catarrhalis BASB019 protein coding sequence #2.
XX
KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
KW upper respiratory tract infection; middle ear infection; therapy; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO9957277-A2.
XX
PD 11-NOV-1999.
XX
PE 03-MAY-1999; 99WO-EP03038.
XX
PR 06-MAY-1998; 98GB-0009683.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX
DR WPI: 2000-062148/05.
XX
PR P-PSDB; AAY55090.
XX
PT Novel BASB019 polynucleotides and polypeptides from Moraxella
PT catarrhalis used to prepare vaccines against bacterial infections -
XX
PS Claim 10; Fig 2; 101pp; English.
XX
CC This sequence encodes a Moraxella catarrhalis BASB019 protein of the
CC invention. The sequences can be used for diagnosis of disease, staging of
CC disease, or determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and for
CC screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB019 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies, and as a

CC target for the screening of antimicrobial drugs. The polypeptides can
CC also be used in vaccine formulations, and to identify agonists and
CC antagonists. The polypeptides, antibodies, agonists and antagonists
CC (which are bacteriostatic) are used for the treatment and prevention of
CC diseases including bacterial infection, otitis media in infants and
CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of in-dwelling devices or by other surgical techniques.
CC The frequency of Moraxella catarrhalis infections has risen dramatically
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BAMB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.

SQ Sequence 519 BP; 139 A; 108 C; 121 G; 151 T; 0 other

Alignment Scores:

Pred. No.:	4,11e-95	length:	519
Score:	855.00	Matches:	171
Percent Similarity:	99.42%	Conservative:	0
Best Local Similarity:	99.42%	Mismatches:	1
Query Match:	99.30%	Indels:	0
DB:	21	Gaps:	0

US-09-674-779B-2 (1-172) x AAZ40352 (1-519)

[illegible]

Qy	161	A1ATPSPeGlnAsnArgArg
Db	481	GCAATGTCAAAATGTCGTG
RESULT 4		
AAZ40353		
ID	AAZ40353	standard; DNA; 519 BP
XX		
AC	AAZ40353;	

XX	01-MAR-2000 (first entry)
DT	
ru	

DE M. catarrhalis BASB019 protein coding sequence #3

KMN BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
 KMN genetic mutation screening; antibody production; vaccine; otitis media;
 KMN bacterial infection; pneumonia; sinusitis; nosocomial infection;
 KMN invasive disease; delayed speech learning; bacteria adhesion prevention
 KMN upper respiratory tract infection; middle ear infection; therapy; ss.

05 Moraxella catarrhalis.

PN WO9957277-A2.

PD 11-NOV-1999.
xy

PF 03-MAY-1999; 99WO-EP03038.
XX

PR 06-MAY-1998; 98GB-0009683.
YY

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XXPI Ruelle J;
YY

DR WPI; 2000-062148/05.
DR P-PSDB: AAY55091.

XX Novel BASR019 pol
PT

PT catarrhalis used to prepare vaccines against bacterial infections

PS Claim 10; Fig 2; 101pp; English

CC This sequence encodes a Moraxella catarrhalis BASB019 protein of the
CC invention. The sequences can be used for diagnosis of disease, staging of
CC disease, or determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and for
CC screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB019 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies, and as a
CC target for the screening of antimicrobial drugs. The polypeptides can
CC also be used in vaccine formulations, and to identify agonists and
CC antagonists. The polypeptides, antibodies, agonists and antagonists
CC (which are bacteriostatic) are used for the treatment and prevention of
CC diseases including bacterial infection, otitis media in infants and
CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of in-dwelling devices or by other surgical techniques.
CC The implantation of Moraxella catarrhalis infections has risen dramatically
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.

Sequence 519 BP; 140 A; 108 C; 120 G; 151 T; 0 other;

Alignment Scores:

Pragmatic Score:	1.68e-94	Length:	519
Pred. No.:	850.00	Matches:	169
Score:	98.84%	Conservative:	1
Percent Similarity:	98.26%	Mismatches:	2
Best Local Similarity:	98.72%	Indels:	0
Query Match:	21	Gaps:	0
DB:			

US-09-674-779B-2 (1-172) X AAZ40353 (1-519)

```

QY 1 MetMetLeuHisIleGlnIleAlaAlaAlaAlaLeuSerValIleuThrPheMet 20
Db 1 ATGATGTTACATATTCAAATTCGCCGCGCGCCCTTATTCGTTACTACTTTATG 60
QY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
Db 61 ACAGGCTGTGCAATTAATCAACAGCAAGCAAGTATGTTCTCTCTAATGACCCACAGGT 120
QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAlaSpasnaSpGluThrValLys 60
Db 121 TACACGTGGGTATCTACTACTGTTGACCTTTGGTACATTAATGATGACTATCAAG 180
QY 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80
Db 181 ACTTTGGCAAGCAGCGTACCAGCTTGGTTATTTGACTTGTGATTCGTATGATTA 240
QY 81 ProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
Db 241 CCGCAAGCTGTGCTGCACTTTCAGCAAGCAACAAATTTTAAACCAATCAACAGCT 300
QY 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
Db 301 CGTGTGTTGGTGCAGGTCTATACCGATGACCGTGGTACGTGATATATATATGTCAGT 360
QY 121 GlyGluArgArgAlaValAlaValAlaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
Db 361 GGGGAAGCGCGTGGCGGTGACCACTATTTGCTTGGTAAAGCATTTATCAAGCC 420
QY 141 SerValGluIleLeuSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGlu 160
Db 421 AGCGTTGAGATTATCACTTTTGGTGAAGAACGCCCTATGCGCATTTGGCAAAATGAAGA 480
QY 161 AlaTyrSerGlnAsnArgArgAlaGluLeuSerTyr 172
Db 481 GCATGCTCACAAATATGCTGCTGCTGAATGCTGTAT 516
RESULT 5
AAZ40354
ID AAZ40354 standard; DNA; 519 BP.
XX
AC AAZ40354;
XX
DT 01-MAR-2000 (first entry)
XX
DE M. catarrhalis BASB019 protein coding sequence #4.
XX
KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
KW upper respiratory tract infection; middle ear infection; therapy; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO9957277-A2.
XX
PD 11-NOV-1999.
XX
PF 03-MAY-1999; 99WO-EP03038.
XX
PR 06-MAY-1998; 98GB-0009683.
XX
PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX
DR WPI: 2000-062148/05.
DR P-PSDB: AAY5092.
XX
PT Novel BASB019 polynucleotides and polypeptides from Moraxella
XX catarrhalis used to prepare vaccines against bacterial infections
XX Claim 10; Fig 2; 101pp; English.

```

```

XX This sequence encodes a Moraxella catarrhalis BASB019 protein of the
CC invention. The sequences can be used for diagnosis of disease, staging of
CC disease, or determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and for
CC screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB019 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies, and as a
CC target for the screening of antimicrobial drugs. The polypeptides can
CC also be used in vaccine formulations, and to identify agonists and
CC antagonists. The polypeptides are used for the treatment and prevention of
CC diseases including bacterial infection, otitis media in infants and
CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of in-dwelling devices or by other surgical techniques.
CC The frequency of Moraxella catarrhalis infections has risen dramatically,
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.
XX
SQ Sequence 519 BP; 137 A; 111 C; 122 G; 149 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 9,06e-94 Length: 519
Score: 844.00 Matches: 169
Percent Similarity: 98.26% Conservative: 0
Best Local Similarity: 98.26% Mismatches: 3
Query Match: 98.03% Indels: 0
DB: 21 Gaps: 0

```

US-09-674-779B-2 (1-172) x AAZ40354 (1-519)

```

QY 1 MetMetLeuHisIleGlnIleAlaAlaAlaAlaLeuSerValIleuThrPheMet 20
Db 1 ATGATGTTACATATTCAAATTCGCCGCGCGCCCTTATTCGTTACTACTTTATG 60
QY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
Db 61 ACAGGCTGTGCAATTAATCAACAGCAAGCAAGTATGTTCTCTCTAATGACCCACAGGT 120
QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAlaSpasnaSpGluThrValLys 60
Db 121 TACGCTGGCGTTACTACACTGCTGTTGCCCTTTGGTACATTAATGATGAGACCGTAA 180
QY 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80
Db 181 GCTTTGGCAAGCAGCTACCCAGTTGTTATTTGACTTGTATCTGTATGATTA 240
QY 81 ProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
Db 241 CCGCAAGCTGTGCTGCACTTTCAGCAAGCAACAAATTTTAAACCAATCAACAGCT 300
QY 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
Db 301 CGTGTGTTGGTGCAGGTCTATACCGATGACCGTGGTACGTGATATATATATGTCAGT 360
QY 121 GlyGluArgArgAlaValAlaValAlaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
Db 361 GGGGAAGCGCGTGGCGGTGACCACTATTTGCTTGGTAAAGCATTTATCAAGCC 420
QY 141 SerValGluIleLeuSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGlu 160
Db 421 AGCGTTGAGATTATCACTTTTGGTGAAGAACGCCCTATGCGCATTTGGCAAAATGAAGA 480

```


OY 161 AlaTrpSerGlnAspArgArgAlaGluLeuSerTyr 172
 Db 481 GCATGGTCACAAATCGTCGTGCAACTGTCTTAT 516

RESULT 6

ID AA006089 standard; DNA; 866 BP.

AC AA006089;

24-JAN-1991 (first entry)

Sequence encoding 16.6kd outer membrane protein (OMP) of H.influenzae.

HI vaccine; OMP; ds.

Haemophilus influenzae.

Key Location/Qualifiers

FT CDS 67..525

FT EP389925-A.

PD 03-OCT-1990.

PF 20-MAR-1990; 90EP-0105205.

PR 29-MAR-1989; 89US-0330229.

PA (UYNE-) STATE UNIV NEW YORK.

PI Murphy TF, Apicella MA;

DR WPI; 1990-298924/40.

DR P-PSDB; AAR07145.

PT - by separation of outer membrane protein of haemophilus influenzae

PT detergent-contg., then detergent-free buffers.

PS Disclosure; Page 8; 22pp; English.

CC Method claimed produces large quantities of the purified OMP, useful

CC in raising antibodies for detection, and as a vaccine against

CC H.influenzae.

XX Sequence 866 BP; 252 A; 168 C; 175 G; 271 T; 0 other;

Alignment Scores:

Pred. No.: 2.85e-23 Length: 866
 Score: 269.50 Matches: 65
 Percent Similarity: 50.00% Conservative: 29
 Best Local Similarity: 34.57% Mismatches: 51
 Query Match: 31.30% Indels: 43
 DB: 11 Gaps: 4

US-09-674-779b-2 (1-172) x AA006089 (1-866)

OY 11 AlaAlaLeuSerValLeu-----ThrPheMetThrGlyCysAlaAsn 25
 Db 13 TCCAGCTTGTCTCCACTTAACCTAAATAAAAACTCATTTAGAGAAATCAATGAC 72
 OY 26 LysSerThrSerGlnValMetValAlaProAsnAlaProThr----- 39
 Db 73 AAATTTGTTAAATCATATTATTAGTGCAGGTTCTGTACTGCATTAGCGGCTGTAGTCC 132
 OY 40 -----GlyTyrThrGlyVal 44
 Db 133 TCTAACACGATGCTGCAGCAGCAATGTCGTCTCAAGTTTGGCGGATACTCT----- 186
 OY 45 TleTyrThrGlyValAlaProLeuValAlaSpasnspluThrValAlaLeuAlaSer 64
 ||||| |||

Db 187 -----GTTGCTGATCTTCACCAACGTTTACAAAC----- 216
 OY 65 LysLeuProSerLeuValTyrPheAspPheAspSerAspGluLeysProGlnAlaAla 84
 Db 217 -----GTAATTTTGGTTTGTATTAATACACATCAACCGGTGAATACGT 261
 OY 85 AlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuVal 104
 Db 262 CAATCTTAGATGCGCAGCGCATATTTAATGCACACGCGAGCTGCTAAGTATAGTA 321
 OY 105 AlaGlyHisThrAspGluArgGlySerArgGlnTyrTrsMetSerLeuGlyGluArg 124
 Db 322 GAAGTATACTGTGAACCGGTACACCAAGATACACATCGCATTAGGACACACGTCT 361
 OY 125 AlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIle 144
 Db 382 GCAGATGCGATTAAAGTATTATTGACAGGTAAAGCTTGTGCTGAATATAGCACA 441
 OY 145 IleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGln 164
 Db 442 GTATCTTACGGTGAAGAAACCTGCAGTATTAGTACGATGACATGAAGCTGATATTCTAA 501

RESULT 7

ID AA003869

AC AA003869;

30-AUG-1990 (first entry)

H.influenzae DNA fragment containing the pOMP-1 gene.

KW outer membrane proteins; pOMP-1; active immunisation;

KW typable H influenzae; non-typable H influenzae; ss.

OS Haemophilus influenzae.

PS Location/Qualifiers

FT CDS 241..702

FT /product=pOMP-1, outer membrane protein of H.influenzae.

PN WO9002557-A.

PD 22-MAR-1990.

PF 31-AUG-1989; 89WO-US03779.

PR 01-SEP-1988; 88US-0239572.

PR 21-AUG-1989; 89US-0396572.

PA (PRAX-) PRAXIS BIOLOGICS IN.

PI Anilionis A, Seid RC, Delch RA, Zlotnick GW, Green BA;

DR WPI; 1990-115815/15.

DR P-PSDB; AAR05797.

PT Outer membrane protein epitopes of Haemophilus influenzae- used in

PT the prodn. of antibodies, in vaccines and for prodn. of reagents for

PS diagnosis.

PS Disclosure; Fig 10; 164pp; English.

CC See also AA003870, R03948 and AAR03949.

XX Sequence 737 BP; 225 A; 134 C; 154 G; 224 T; 0 other;

SO Alignment Scores:

Pred. No.: 1.07e-22
 Score: 264.00
 Percent Similarity: 68.93%
 Best Local Similarity: 49.51%
 Query Match: 30.66%
 DB: 11
 Matches: 51
 Conservative: 20
 Mismatches: 32
 Indels: 0
 Gaps: 0

US-09-674-779B-2 (1-172) x AA003869 (1-737)

QY 70 VALTYRPhaspPheaspSeraspGluilelyspProGlnAlaAlaAlaileuaspGlu 89
 |||||
 DB 391 GTTTATTTCGCTTTTGAATAAATATGACATTACGCTGTAATCGTTCAATCTTAGATCGC 450
 QY 90 GlnAlaGlnPheleuThrThrasnGlnThrAlaArgValleuValAlaGlyHisThrAsp 109
 |||||
 DB 451 CACGCTGCATATTAAATGACACACACACACCTGCTAAGATTAGTAGAAGGTACACGCTAT 510
 QY 110 GluArgGlySerArgGluTyrAsnMetSerleuGlyGluArgAlaValAlaValArg 129
 |||||
 DB 511 GAACCTGCTACACGACATACACATCCATTAGCCCAACGCTGCGATGACGATTAA 570
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluilelieserPheGlyGlu 149
 |||||
 DB 571 GGTATTATTCGCTGTAAGAGTGTGATGCTGTAATTAAGGACAGATCTTAGCGTAA 630
 QY 150 GluArgProIleAlaPheGlyThrAsnGlnGluAlaTyrSerGlnAsnArgArgAlaGlu 169
 |||||
 DB 631 GAAAACCTGCAGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 690
 QY 170 LeuSerTyr 172
 |||||
 DB 691 TTAGCGTAC 699

RESULT 8
 ID AAN81194 standard; DNA; 867 BP.
 AC AAN81194;
 DT 10-OCT-1990 (first entry)
 DE 16600 dalton outer membrane protein of non-typable H influenzae.
 KW Haemophilus influenzae; 16600 dalton outer membrane protein (OMP);
 KW pneumonia; meningitis; ss
 KW Haemophilus influenzae.
 OS
 XX
 FH Key location/Qualifiers
 FT CDS 68..530
 FT /*tag= a
 FT /product=16600 dalton OMP
 XX
 XX EP281673-A.
 XX PD 14-SEP-1988.
 XX PF 12-NOV-1987; 87EP-0116699.
 XX PR 08-OCT-1987; 87US-0092948.
 XX PR 08-NOV-1986; 86US-0932872.
 XX PA (UYNV-) RES UNIV NEW YORK.
 XX PI Murphy TF, Apicella MA;
 XX WPI: 1988-258472/37.
 XX P-PSDB; AAP80593.
 XX
 XX Plasmids, E.coli, hybridomas and antibodies from H. influenzae protein -
 XX for use as vaccine and detection of Haemophilus influenzae
 XX Disclosure; ; pp; English.

XX An epitope of the protein encoded by this sequence is recognised by
 CC the 7F3 antibody in a number of isolates. Monoclonal antibodies to
 CC the 16600 D OMP are produced by hybridomas and used to detect
 CC H. influenzae. A suspect sample is contacted with the Mab in the
 CC presence of an indicator such as radiolabelled probe.
 CC
 XX

SO Sequence 867 BP; 253 A; 169 C; 174 G; 271 T; 0 other;

Alignment Scores:

Pred. No.: 1.34e-22
 Score: 264.00
 Percent Similarity: 68.93%
 Best Local Similarity: 49.51%
 Query Match: 30.66%
 DB: 9
 Matches: 867
 Conservative: 51
 Mismatches: 20
 Indels: 32
 Gaps: 0

US-09-674-779B-2 (1-172) x AAN81194 (1-867)

QY 70 VALTYRPhaspPheaspSeraspGluilelyspProGlnAlaAlaAlaileuaspGlu 89
 |||||
 DB 218 GTATTATTTCGCTTTTGAATAAATATGACATCCGCTGTAATCGTTCAATCTTAGATCGC 277
 QY 90 GlnAlaGlnPheleuThrThrasnGlnThrAlaArgValleuValAlaGlyHisThrAsp 109
 |||||
 DB 278 CACGACGATATTAAATGACACACGCTGCTAAGATTAGTAGAAGGTAAATCTGAT 337
 QY 110 GluArgGlySerArgGluTyrAsnMetSerleuGlyGluArgAlaValAlaValArg 129
 |||||
 DB 338 GAACGCTGTAACACGACATACACATCCATTAGCCCAACGCTGCGATGACGATTAA 397
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluilelieserPheGlyGlu 149
 |||||
 DB 398 GGTATTATTCGCTGTAAGAGTGTGATGCTGTAATTAAGGACAGATCTTAGCGTAA 457
 QY 150 GluArgProIleAlaPheGlyThrAsnGlnGluAlaTyrSerGlnAsnArgArgAlaGlu 169
 |||||
 DB 458 GAAAACCTGCAGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 517
 QY 170 LeuSerTyr 172
 |||||
 DB 518 TTAGCGTAC 526

RESULT 9
 ID AAQ45440 standard; DNA; 1019 BP.
 AC AAQ45440;
 DT 18-OCT-1994 (first entry)
 DE Fragment encoding outer membrane protein P6 of H. influenzae.
 KW Outer membrane protein; vaccine; antibody; treatment; detection;
 KW identification; pneumonia; bacteraemia; meningitis;
 KW postpartum sepsis; acute febrile tracheobronchitis; neonatal sepsis;
 KW acute otitis media; Haemophilus influenzae; ds.
 KW Haemophilus influenzae.
 OS
 XX
 FH Key location/Qualifiers
 FT CDS 220..681
 FT /*tag= a
 FT /product= P6 outer membrane protein
 XX
 XX US5300632-A.
 XX PD 05-APR-1994.
 XX PF 18-NOV-1986; 86US-0932872.
 XX PR 18-NOV-1986; 86US-0932872.
 XX PR 08-OCT-1987; 87US-0092948.

PR 29-MAR-1989; 89US-0330229.
 PR 12-DEC-1991; 91US-0807049.
 XX (UYNV) UNIV NEW YORK STATE RES FOUND.
 PA Apicella MA, Murphy TF;
 PI WPI. 1994-109467/13.
 XX P-PSDB; AAR51161.
 DR
 XX purifying outer membrane protein of Haemophilus influenzae - by
 PT extr. with detergent buffer, treatment with RNase and
 PT solubilising by heat-treating in detergent-free buffer
 XX
 PS Example 4; Column 21-22; 24pp; English.
 XX
 CC The P6 outer membrane protein (OMP) is conserved among strains of
 CC Haemophilus influenzae and so antibodies to the OMP can be used to
 CC identify H. influenzae in clinical samples. The OMP can also be
 CC used in the development of vaccines against H. influenzae which
 CC causes pneumonia, bacteraemia, meningitis, postpartum sepsis and
 CC acute febrile tracheobronchitis in adults and neonatal sepsis and
 CC acute otitis media in infants and children.
 XX
 SQ Sequence 1019 BP; 310 A; 200 C; 186 G; 323 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.68e-22 Length: 1019
 Score: 264.00 Matches: 51
 Percent Similarity: 68.93% Conservative: 30
 Best Local Similarity: 49.51% Mismatches: 32
 Query Match: 30.66% Indels: 0
 DB: Gaps: 0
 US-09-674-779B-2 (1-172) x AAO45440 (1-1019)
 QY 70 ValTyrPheASPheASPserASPgluIleLysProGlnAlaAlaIleLeuaspGlu 89
 DB 370 GATATTGTTGGTTTGAATAATACGATACACCGGTGAATACGTTCAATCTTAGATCGG 429
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
 DB 430 CACGCGAGCATATTAAATGCAACGCCAGCTGCTAAAGTATTAGTAGAAGATATCAT 489
 QY 110 GlnArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 DB 490 GAACGTGTACACCAAGATACACATCCGATTAGACACACGTGCGAGATGCAATTAA 549
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
 DB 550 GGTATTATTAGCAGTAAGGTGTGATGCTGTAATAATAGGACACATCTTAGCGGTAA 609
 QY 150 GlnArgProIleAlaPheGlyThrAsnGlnGluAlaIleTyrSerGlnAsnArgAlaGlu 169
 DB 610 GAAAAACCTGCAGATTAGGTACAGATGAGCTGCATATTCTTAAAAACCGTCGTGCA 669
 QY 170 LeuSerTyr 172
 DB 670 TTACGCTAC 678
 RESULT 10
 AAT42063/c
 ID AAT42063 standard; DNA; 1830121 BP.
 XX
 AC AAT42063;
 XX
 DT 14-SEP-1999 (first entry)
 XX
 XX Haemophilus influenzae complete genome sequence.
 DE
 XX Genome: bacterium; Haemophilus influenzae; computer readable medium;
 KW expression modulating fragment; regulation; gene expression; vector;
 KW organism; open reading frame; ORF; ds.

XX
 OS Haemophilus influenzae.
 XX
 PN WO9633276-A1.
 XX
 PD 24-OCT-1996.
 XX
 PF 22-APR-1996; 96WO-US053320.
 XX
 PR 07-JUN-1995; 95US-0487429.
 PR 21-APR-1995; 95US-0426787.
 PR 07-JUN-1995; 95US-0476102.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX (UYNV) UNIV JOHNS HOPKINS.
 PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
 XX WPI; 1996-485782/48.
 DR
 XX Haemophilus influenzae Rd genome recorded on computer readable
 PT medium - useful for identifying commercially important nucleic acid
 PT fragments by homology searching
 XX
 PS Claim 1; Page 77.2-77.1091; 1291pp; English.
 XX
 CC This sequence represents the complete genome sequence of the bacterium
 CC Haemophilus influenzae strain Rd. The invention relates to a computer
 CC readable medium (CRM) having recorded upon it the complete H. influenzae
 CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
 CC sequence at least 99% identical to (1). By providing the full-length
 CC genomic sequence in a computer readable form, it is possible to identify
 CC commercially important nucleic acid fragments and expression modulating
 CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
 CC regulate the expression of a nucleic acid molecule. Vectors and altered
 CC organisms comprising the predicted ORFs can be used to produce any of the
 CC polypeptide fragments of the H. influenzae Rd genome.
 XX
 SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;
 Alignment Scores:
 Pred. No.: 5.67e-18 Length: 1830121
 Score: 264.00 Matches: 51
 Percent Similarity: 68.93% Conservative: 20
 Best Local Similarity: 49.51% Mismatches: 32
 Query Match: 30.66% Indels: 0
 DB: Gaps: 0
 US-09-674-779B-2 (1-172) x AAT42063 (1-1830121)
 QY 70 ValTyrPheASPheASPserASPgluIleLysProGlnAlaAlaIleLeuaspGlu 89
 DB 401840 GTTATTATTCGTTTGTAAATATGACATTACTGTGAATACGTTCAATCTTAGCGCG 401781
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
 DB 401780 CACGCTCATATTAAATGCAACGCCAGCTGCTAAAGTATTAGTAGAAGATATCAT 401721
 QY 110 GlnArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 DB 401720 GAACGTGTACACCAAGATACACATCCGATTAGACACACGTGCGAGATGCAATTAA 401661
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
 DB 401660 GGTATTATTAGCTGTAAGGTGTGATGCTGTAATAATAGGACACATCTTAGCGGTAA 401601
 QY 150 GlnArgProIleAlaPheGlyThrAsnGlnGluAlaIleTyrSerGlnAsnArgAlaGlu 169
 DB 401600 GAAAAACCTGCAGATTAGGTACAGATGAGCTGCATATTCTTAAAAACCGTCGTGCA 401541
 QY 170 LeuSerTyr 172
 DB 401540 TTACGCTAC 401532

RESULT 11
 AAN80226
 ID AAN80226 standard; DNA; 737 BP.
 AC AAN80226;
 XX
 DT 12-JAN-1991 (first entry)
 XX
 DE Sequence of Haemophilus influenzae Praxis Biologics Outer
 DE Membrane Protein-1 (PBOM-1) gene.
 XX
 KW Vaccine; diagnosis; epitope; passive immunisation; ss.
 OS Haemophilus influenzae.
 XX
 FH Key Location/Qualifiers
 FT CDS 241..702
 FT /*tag= a
 XX
 PN MO804932-A.
 XX
 PD 14-JUL-1988.
 XX
 PE 23-DEC-1987; 87WO-US03423.
 XX
 PR 11-DEC-1987; 87US-0132073.
 PR 02-MAR-1987; 87US-0020849.
 PR 31-DEC-1986; 86US-0948364.
 XX
 PA (PRAX-) PRAXIS BIOLOGICS IN.
 XX
 PI Deich RA, Zlotnick G, Green B;
 XX
 DR WPI; 1988-205305/29.
 DR P-PSDB; AAP80665.
 XX
 PT Pure peptide related to epitope of Haemophilus influenzae -
 PT used as immunogens in vaccines and for producing antibodies for
 PT passive immunisation and assays
 XX
 PS Disclosure; Fig 10; 129pp; English.
 XX
 CC A pure antigenic peptide or protein related to an epitope of
 CC Haemophilus influenzae is claimed. Also claimed is a recombinant
 CC vector comprising a DNA sequence coding for an antigenic
 CC determinant of an Haemophilus influenzae outer membrane protein,
 CC the transformed cell, a subunit vaccine in a pharmaceutical carrier,
 CC a method of immunising humans and an assay for Haemophilus influenzae.
 XX
 SQ Sequence 737 BP; 225 A; 133 C; 154 G; 225 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,36e-22 Length: 737
 Score: 259.00 Matches: 50
 Percent Similarity: 67.968 Conservative: 20
 Best Local Similarity: 48.548 Mismatches: 33
 Query Match: 30.088 Indels: 0
 Gaps: 0
 DB: 9
 US-09-674-779b-2 (1-172) x AAN80226 (1-737)
 QY 70 valtyrphaspheaspsersaspgtulleysproginalaialaleuaspclu 89
 Db 391 gtttatttggctttgataaataatgacattactggtgatacgttcaatcttaacgog 450
 QY 90 glnalaglnphelenthrhrasnnglnthrralaagvalaaleuvalaiglyhisthrsp 109
 Db 451 cacgctgcattatttaaaccaacacacgctgctaaagattagtagaaggtaacactgat 510
 QY 110 gtuatrglyserargglutyrasnmetserleuglygluarargalavalalvalary 129
 Db 511 gaacggtgtacacacgaatatacaacatcgattagcccaacgctgacagatgcattaaa 570

QY 130 AsnTyrlenuleuglylysclylleasnnglnalaServalGluIleIleSerphleglylu 149
 Db 571 ggttatttggctgtgataaagctgtgataatgacacagatattcttaacggtcaa 630
 QY 150 gluargprollaalealaphleglythrasnnglnualatpserclnasargaralalu 169
 Db 631 gaaaaacctgcagattttaggtcattgataatctgatttcaaaaacccgtgctgacgtg 690
 QY 170 leuSerTyr 172
 Db 691 ttacgctac 699
 RESULT 12
 AAT67789
 ID AAT67789 standard; DNA; 540 BP.
 AC AAT67789;
 XX
 DT 29-JUL-1997 (first entry)
 XX
 DE H. pylori outer membrane protein ORF 31262.aa.
 XX
 KW Outer membrane; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis; ds.
 XX
 OS Helicobacter pylori.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..540
 FT /*tag= a
 XX
 PN W09640893-AL.
 XX
 PD 19-DEC-1996.
 XX
 PE 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTRA) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Meligaerd BL;
 XX
 DR WPI; 1997-052306/05.
 DR P-PSDB; AAM20394.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 Claim 1; Page -: 1481pp; English.
 This sequence encodes a H. pylori outer membrane protein.
 The protein may be used in a vaccine to prevent or treat H. pylori
 infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 CC Note: This DNA sequence is not reproduced in the specification and
 CC has been derived from the related specification, W09719098.

SQ Sequence 540 BP; 190 A; 88 C; 128 G; 134 T; 0 other;

Alignment Scores:

Pred. No.:	5.83e-17	Length:	54
Score:	215.50	Matches:	37
Percent Similarity:	50.00%	Conservative:	50
Best Local Similarity:	32.76%	Mismatches:	66
Query Match:	25.03%	Indels:	21

US-09-674-779B-2 (1-172) x AAT67789 (1-540)

OY		15	SerValLeuThrPheMet-----ThrGlyCysAlaAsnIlys-----	26
Dd		13	ICTGTATTATTTAGTTCCTGGTAGCTTTTTTATTGGTAGTGCGCTAGCATATAAATGGAT	72
OY		27	-----SerThrSerGlnValMetValAlaPro-----Asn	36
Dd		73	AATAAGACTGTGCCTGCCGATGTGACACTAAAGCGGTTCCAGACTGCGCGCTGTACTACA	132
OY		37	AlaProThrGlyLysTrhGlyValIleTyTrhGlyValAlaProLeuValAspAsnAsp	56
Dd		133	GAAACCAGCTCCAGAGAAGAAGACGCTTAACAAGAGCCAGCTCCAGTGGTTGAAGAAAAG	192
OY		57	GluTrhValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSer	76
Dd		193	CCGGCTATTGAA-----AGCGGAGCACTATCATCGCTCTATTATTATTGATTGTGACAAG	246
OY		77	AspGluIleLysProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThr	96
Dd		247	TATGAGTCTCAAATAATCCGATCAAGACGACTTAGATGAGATGCTGCCAAAANAAGTAAGAA	306
OY		97	AsnGlnThrAlaArgValLeuValAlaGlnIstHisArgLysGlySerArgLysTyr	116
Dd		307	AACCAC--ATGCAAGTGGCTTTGGAAAGCCAAATACCAGATGATTTGGCTTACGGAATAC	363
OY		117	AsmMetSerLeuGlyLysLysArgValAlaValAlaValArgAsnTyrLeuLeuGlyLysGly	138
Dd		364	AACCAGACGGTGTGGCGCTTAAAAAGCACTTTGACCGTGAAAAACGCTTAGTATTAAGAGG	422
OY		137	IleAsnIlnAlaSerValGluIleLeuSerPheGlyGluGlyArgProIleAlaPheGly	156
Dd		424	GTAACAAAAGATGTGATCAAAACCATCACGTTTGGGGAAGAACCAAMCCCAATGGGTCOA	483
OY		157	ThrAsnGluGluAlaTrpSerGlnAsnArgValGlyLeu	170
Dd		484	AAAATCTAGAAATGTTTACAGAGAAACCAAGAGTGGATGTC	525

RESULT 13

ID AAT77469 standard; DNA; 540 BP

AC AAT77469;

DT 11-AUG-1997 (first entry)

DE H. pylori outer membrane protein ORF 31262.aa.

KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease
 KW activator; inhibitor; bacterial life cycle; vaccine; immune;
 KW detection; antisense; inhibition; ds.

OS Helicobacter pylori.

	Location/Qualifiers
FH	Key
FT	1..540
FT	/*tag= a
FT	

PN WO9719098-A1.

PD 29-MAY-1997

XX

PF	15-NOV-1996:	96MO-US18542.
xx		
PR	17-NOV-1995:	95US-0561469.
xx		
PA	(ASTR) ASTRA AB.	
xx		
P1	Smith DH:	
xx		
DR	WPI: 1997-298052/27.	
xx		
DR	P-PSDB; AAM24651.	
xx		
PT	Helicobacter pylori nucleic acid sequences and related proteins -	
PT	used for diagnostics and therapeutics	
xx		
PS	Claim 1; Page 108; 235pp; English.	
xx		
CC	This sequence encodes an H. pylori outer membrane protein.	
CC	Helicobacter pylori has been strongly linked to chronic gastritis and	
CC	duodenal ulcer disease. The nucleic acid sequences of the invention	
CC	are used to evaluate compounds, especially activators or inhibitors of	
CC	bacterial life cycle, for the ability to bind an H. pylori nucleic acid	
CC	sequence. The nucleic acid sequences, and corresponding proteins, are	
CC	also useful for generating vaccines for immunising subjects against H.	
CC	pylori or for use in detecting the presence of Helicobacter species in	
CC	a sample. Antisense nucleic acid sequences of these sequences are	
CC	used to inhibit expression of a gene from Helicobacter species. H.	
CC	pylori whole genomic DNA was isolated and nebulised to a median size of	
CC	2000 bp. Purified DNA fragments were blunt-ended and ligated to unique	
CC	BstXI-linker adaptors in 100-1000 fold molar excess. These linkers are	
CC	complementary to the BstXI-cut pMP vectors, while the overhang is not	
CC	self-complementary. Therefore the linkers will not concatenate inserts	
CC	will the cut vector re-ligate itself easily. The linker-adaptor inserts	
CC	are ligated to each of the 20 pMP vectors to construct a series of	
CC	shotgun subclone libraries. The purified DNA samples were then	
CC	sequenced.	
CC	Note: The ORF/protein reference number for this sequence was obtained	
CC	from the related specification, W09640893.	
xx		
SQ	Sequence 540 BP; 190 A; 88 C; 128 G; 134 T; 0 other;	
	Alignment Scores:	
	Pred. No.:	5.83e-17
	Score:	215.50
	Percent Similarity:	50.00%
	Best Local Similarity:	32.76%
	Query Match:	25.03%
	DB:	18
		Gaps: 5
US-09-674-779B-2 (1-172) x AAT77469 (1-540)		
QY	15 SerValLeuThrPheMet-----ThnGlyCysAlaAsnIys-----	26
	: :	:
DB	13 TCTGTATTTAGTTTCTTGGTAGCTTTTATTGGTAGTGGCTAGACATAAATGAT	72
QY	27 -----SerThrSerGlnValMetValAlaPro-----Asn	36
DB	73 AATAAGACTGCGGTGCGATGTGACACACTAAAGCGGTGACAGCTGCGCTGACTACA	132
QY	37 AlaProThrGlyTyrThrGlyValIleIleTyrThrGlyValAlaProLeuValAspAsn	56
	: :	:
DB	133 GAACCAGCTCCAGAGAAAGAACGCTTAACAAGAACCCAGCTCCAGGTGGTGAAGAAAG	192
QY	57 GlnThrValYsAlaLeuAlaSerIysLeuProSerLeuValTyrPheAspPheAspSer	76
	: :	:
DB	193 CCGGCTATTGTA-----ACGGAGCACTACGCTCTTATTTATTTGATTGGACAAAG	246
QY	77 AspGlnIleIysProGlnAlaIleAlaIleLeuAspGlnAlaGlnPheLeuThrThr	96
	: :	:
DB	247 TATGAGATCAAAAGAAATCCGATCAAGAGACTTTAGATGAGATGTCGCAAAAGCTAAAGAA	306
QY	97 AsnGlnThrAlaIleGValAlaGValAlaGlnIleThrAspGlnIleArgGlySerArgGlnTyr	116
	: :	:
DB	307 AACCA---ATGCAAGTGCCTTTTGAAGGAATACCGATGGAATTTGGCTAGAGGAATAC	363

US-09-674-779B-2' (1-172) x AAT77469 (1-540)

```

QY      13 SerValLeuThrPheMet-----ThrGlyCysAlaAsnLys----- 26
      |||||  ::|||:::
Db      13 TCTATTATTAGTTCTTGGTAGCTTTTATTGTAGTAGTGTGAGTCATTAATAATGAT 72
QY      27 -----SerThrSerGlnValMetValAlaPro-----Asn 36
      |||||  |||  |||||
Db      73 AATAAGACTGTGGCTGGCCATGTGACACATTAAGCGGTTCCAGACTGCCCGTTACTACCA 132
QY      37 AlaProThrGlyTyrThrGlyValIleIleYrThrGlyValAlaProLeuValAspAsnAsp 56
      |||  |||||:::|||||
Db      133 GAACCAAGCTCCAGGAAGAAGAACCTTAACAAGACCAGCTCCAGCTGGTTGAGAAAG 192
QY      57 GluThrValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSer 76
      ::::  ::::  ::::  ::::  ::::  ::::
Db      193 CCGCGCTATTGTA-----ACGGGAGCTATCAACGCTTCATTATTATTGATTGTGACAA 248
QY      77 AspGluIleLysProGlnAlaIleAlaIleLeuAspGluGlnAlaGlnPheLeuThrThr 96
      |||||  |||||  |||||  |||
Db      247 TATGAGATCAAAAGATCCGATCAAGAGACTTTAGATGAGATGTGCGAAAAGCTTAAGAA 300
QY      97 AsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluIleArgGluIleThr 116
      |||  ::|||:::  |||  |||||  |||||  |||||
Db      307 AACCA---ATGCAAGTCTTTTGGAAAGCAATCCGATGAATTTGGCTACAGGAATAC 363

```

Alignment Scores:		Pred. No.:		5.83e-17		Length:	
Score:		215.50		Matches:		540	
Percent Similarity:		51.72%		Conservative:		34	
Best Local Similarity:		32.18%		Mismatches:		63	
Query Match:		25.03%		Indels:		21	
DB:		22		Gaps:		5	
US-09-674-779B-2 (1-172) x AAF25593 (1-540)							
QY	15	SerValleuThrPheMet-----	-----	ThrGlyCysAlaAsnLysSerThr	28		
DB	13	TCTGATTTAGTTCTTGTAAGCTTTTATGTGTAAGCTGCGTGTGACATATAAATGAT	72				
QY	29	SerGlnValMetValAlaPro--AsnAlaProThrGlyTyrThrGlyValIleTyrThr	47				
DB	73	AATAAGACTGTGGCGGCGCATGTGATGCTTAAACGGTTCAACAGTCACTGTTACTACA	132				
QY	48	-----	-----	GlyValAlaProLeuValAspAsnAsp	56		
DB	133	GAACCACTCCAGAAAGAAAGAGCCTTAACAAAGACCAAGCTCCAGTGTGAAGAAAA	192				
QY	57	GluThrValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspAspSer	76				
DB	193	CCGCGCTGTGAG-----	AGCGGCACTATCATCGCTCTCATTTATTATTGATTTTGACACAG	248			
QY	77	AspGluLeuLysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThr	96				
DB	247	TATGATAATCAAAATCCGATCAAGACCTTATGATGAGTCGTCANAAAGCTAAAGAA	306				
QY	97	AsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluAspGlySerArgGluTyr	116				
DB	307	AACAC---ATGCAAGTGTGTTGGCAAGCAATACCGATGAAATTTGGCTTACGCAATAC	363				
QY	117	AsnMetSerLeuLysGluArgAlaValAlaValaLysAsnTyrLeuLeuGlyLysGly	136				
DB	364	AACCAACGCGTGTGCGCTTAAAGAGACTTGAGCGTGAAAAACGCTTATGATTTAAAGG	423				
QY	137	IleAsnGlnAlaSerValGluIleIleSerPheGlyGluArgProIleAlaPheGly	156				
DB	424	GTAAGAAAAAGATATGATCAAAACCATACGATTTGGTGAACCAACCAATGGGCCAA	483				
QY	157	ThrAsnGluGluArgIlePheSerGlnAsnArgArgAlaGluLeu	170				
DB	484	AAAACTRAGAGATGTTATTAAGAAAAAGAGAGTGATGTC	525				
RESULT 15							
ID	AAT68048	AAT68048 standard; DNA; 561 BP.					
AC	AAT68048;						
XX	21-JUL-1997	(first entry)					
DE	H. pylori outer membrane protein ORF 07gp3516orf4.						
XX							
KW	Cytoplasmic; vaccine; prevention; treatment; infection; identification;						
KW	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;						
XX	duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.						
OS	Helicobacter pylori.						
XX							
FT	Key	Location/Qualifiers					
FT	CDS	1..561					
FT		/*tag= a					
XX		/note= "no stop codon given"					
PN	WO9640893-A1.						
PD	19-DEC-1996.						
PF	06-JUN-1996; 96WO-US09122.						

Db 100 GATTAATTAAGACTGTGGCTGGCGGATGTAGCGCTAAACGGTTACAGTGCACCTGTTACT 159
 QY 47 ThrglyValAlaProLeuValAspAsnspglunhrVallysala---LeuAlaSerLys 65
 Db 160 ACAGAACACAGCTCCGAGAAAGAAAGAGCTTAAACAGAGCCAGCTCCAGTGGTTGAAGAA 219
 QY 66 LeuProSerLeu-----ValTyrPheAspPheAspSerAsp 77
 Db 220 AAGCGCGCTATTGAAAGCGGAGCTATCATCCCTCTATTATTATTGATTGACAACTAT 279
 QY 78 GluIleLysProGlnAlaAlaAlaIleLeuAspLuglnAlaGlnPheLeuThrThrAsn 97
 Db 280 GAGATCAAGAAAGATCCGATCAAGAGACTTTAGATAGATCCGTCAAAAAGCTAAAGAAAGAA 339
 QY 98 GlnThrAlaArgValLeuValAlaGlyHisThrAspLuglnArgLysSerAlaGlyLysAsn 117
 Db 340 CAC---ATGCAGAGTCTTTTGGAAAGCCAAATCCGATTAATTTGGCTTAGCGAATACAC 396
 QY 118 MetSerLeuGlyLuglnArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyLe 137
 Db 397 CAAGCGCTTGGCGTTAAAGAGACTTGGAGCGTGAAGAAACGCTTAGTCATTTAAAGGGGTA 456
 QY 138 AsnGlnAlaSerValGluIleIleSerPheGlyLuglnArgProIleAlaPheGlyThr 157
 Db 457 GAAAAAGATATGATCAAAACCATCATCGTTTGGTGAACCAACCAATGCGCCCAAAA 516
 QY 158 AsnGlnLuglnAlaTrpSerGlnAsnArgArgAlaGluLeu 170
 Db 517 ACTAGAGATGTTTACAAAGAAACAGAGAGTGATGTC 555
 RESULT 17
 AAV90843
 ID AAV90843 standard; DNA; 1236 BP.
 AC AAV90843;
 XX 18-FEB-1999 (first entry)
 DT
 XX Nucleotide sequence of cluster 2.
 DE
 XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
 OS Helicobacter pylori.
 XX
 PN WO9849314-A2.
 XX
 PD 05-NOV-1998.
 XX
 PF 27-APR-1998; 98WO-US08487.
 XX
 PR 14-OCT-1997; 97US-0061958.
 PR 25-APR-1997; 97US-0045107.
 XX
 XX (GENE-) GENELABS TECHNOLOGIES INC.
 PA
 PI Chow TP, Fry KE, Lim MY, McAtee CP;
 PI WPI; 1999-009433/01.
 DR
 XX New Helicobacter pylori antigens and related nucleic acid sequences
 PT - useful in serological diagnosis and protective vaccines, providing
 PT long-lasting immune response
 XX
 PS Claim 20; Page 281; 402pp; English.
 XX
 CC The present sequence encodes a Helicobacter pylori antigenic protein
 CC that is characterised by immunoreactivity with H. pylori positive
 CC antisera. The proteins are highly immunogenic and induce a long-lasting
 CC immune response that persists even after antimicrobial treatment. In
 CC antibody detection assays, on sera, plasma, urine, saliva etc., they are
 CC highly sensitive and specific. The specification also describes 69
 CC previously unrecognised immunogenic cluster families. H. pylori antigens

CC are used to detect H. pylori-specific antibodies, for diagnosing
 CC infection or to confirm eradication of infection, and in vaccines to
 CC protect against H. pylori infection and related diseases (gastritis,
 CC peptic ulcer, gastric adenocarcinoma/lymphoma).
 XX
 SO Sequence 1236 BP; 430 A; 220 C; 253 G; 332 T; 1 other;
 Alignment Scores:
 Pred. No.: 5,69e-16 Length: 1236
 Score: 211.50 Matches: 54
 Percent Similarity: 52.02% Conservative: 36
 Best Local Similarity: 31.21% Mismatches: 72
 Query Match: 24,568 Indels: 11
 DB: 20 Gaps: 4
 US-09-674-779b-2 (1-172) x AAV90843 (1-1236)
 QY 8 AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLysSer 27
 Db 297 TCTTGTGCAATTAGTGTCTTGGTACCTTTTATTTGATGAGTGGCTGATGATATAATG 356
 QY 28 ThrSerGlnValMetValAlaPro---AsnAlaProThrGlyTyrThrGlyValIleYr 46
 Db 357 GATATATAGACGTGTGGCTGGCATGTGACCGCTAAACGTTCAAGCTGCACCTGTTACT 416
 QY 47 ThrglyValAlaProLeuValAspAsnspglunhrVallysala---LeuAlaSerLys 65
 Db 417 ACAGAACACAGCTCCAGAGAAAGAGAGCGCTAAACAGCCAGCTCCAGTGGTTGAAGAA 476
 QY 66 LeuProSerLeu-----ValTyrPheAspPheAspSerAsp 77
 Db 477 AAGCGCGCTATTGAAAGCGGAGCTATCATCCCTCTATTATTATTGATTGACAACTAT 536
 QY 78 GluIleLysProGlnAlaAlaAlaIleLeuAspLuglnAlaGlnPheLeuThrThrAsn 97
 Db 537 GAATCAAGAAATCCGATCAAGAGACTTATGATGAGATCGTCAAAAAGCTAAAGCAAAAC 596
 QY 98 GlnThrAlaArgValLeuValAlaGlyHisThrAspLuglnArgLysSerArgLysGlyLeu 117
 Db 597 CAC---ATGCAGAGTCTTTTGGAGCCAAATACCGATGAATTTGGCTTAGCGAATACAC 653
 QY 118 MetSerLeuGlyLuglnArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyLe 137
 Db 654 CAAGCGCTTGGCGTTAAAGAGCTTGGAGCGGAAAGAGCGCTTAGTCATTTAAAGGGGTA 713
 QY 138 AsnGlnAlaSerValGluIleIleSerPheGlyLuglnArgProIleAlaPheGlyThr 157
 Db 714 GAAAAAGATATGATCAAAACCATCATCGTTTGGTGAACCAACCAATGCGCCCAAAA 773
 QY 158 AsnGlnLuglnAlaTrpSerGlnAsnArgArgAlaGluLeu 170
 Db 774 ACTAGAGATGTTTACAAAGAAACAGAGAGTGATGTC 812
 RESULT 18
 AAV90545
 ID AAV90545 standard; DNA; 1610 BP.
 AC AAV90545;
 XX 18-FEB-1999 (first entry)
 DT
 XX Nucleotide sequence from clone Gia.
 DE
 XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
 OS Helicobacter pylori.
 XX
 PN WO9849314-A2.
 XX
 PD 05-NOV-1998.
 XX
 PF 27-APR-1998; 98WO-US08487.

US-09-674-779B-2 (1-172) x AAT74194 (1-435)

QY 50 AlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSerLeu 69
 Db 67 GCTCCAGGAGGTTGAGAAAGCCGCGTGTGAG-----AGCGGACTATCATCGCTTC 120
 QY 70 ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGlu 89
 Db 121 ATTATATTTGATTTTGACAGTATGCAATCAACCAATCCATCAAGAGACTTTAGATGAG 180
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
 Db 181 ATCGTGGCAAAAAGCTAAAGAAACCCAC---ATCGAAGTGGCTTTGGAAAGCAATACCGAT 237
 QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyLysArgArgAlaValAlaArg 129
 Db 238 GAATTTGGCTCTTACGAAATTCACCAAGCCGCTGGCTTAAAGAGACTTTGACCGTGA 297
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyLys 149
 Db 298 AACGCTTACGTTATTTAAAGGGTAGAAAGATATGATCAAAACCATCACTTTTGGTGA 357
 QY 150 GlnArgProIleAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArgArgAlaGlu 169
 Db 358 ACCAAACCCAAATGACCCCAAAAACCTAGAGAGTGTATTAAGAAACAGAGAGTGAT 417
 QY 170 Leu 170
 Db 418 GTC 420

RESULT 20
 ABL91202
 ID ABL91202 standard; DNA; 579 BP.
 AC ABL91202:
 DT 29-JUL-2002 (first entry)
 XX Chlamydia pneumoniae cp7090 ORF DNA, SEQ ID NO:38.
 DE Chlamydia pneumoniae cp7090 ORF DNA, SEQ ID NO:38.
 XX Chlamydia pneumoniae.
 OS Chlamydia pneumoniae.
 KM Chlamydia pneumoniae; antigen; immunogen; vaccine; diagnosis;
 KW human respiratory disease; cardiovascular disease; atherosclerosis;
 KW coronary artery disease; carotid artery stenosis; myocardial infarction;
 KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
 KW strain CWL029; open reading frame; ORF; gene; ds.
 XX Chlamydia pneumoniae.
 XX Key Location/Qualifiers
 FH CDS 1..579
 FT /tag- a
 FT /product- "cp7090"
 FT sig_peptide 1..60
 FT /tag- b
 FT mat_peptide 61..576
 FT /tag- c
 FT /product- "Mature protein"
 XX
 XX PD 10-JAN-2002.
 PN WO200202606-A2.
 XX
 XX PF 03-JUL-2001; 2001MO-IB01445.
 PR 03-JUL-2000; 2000GB-0016363.
 PR 11-JUL-2000; 2000GB-0017047.
 PR 21-JUL-2000; 2000GB-0017983.
 PR 07-AUG-2000; 2000GB-0019368.
 PR 18-AUG-2000; 2000GB-0020440.
 PR 14-SEP-2000; 2000GB-0022583.
 PR 10-NOV-2000; 2000GB-0027549.
 PR 22-DEC-2000; 2000GB-0031706.

XX (CHIR-) CHIRON SPA.
 PA Rattl G, Grandi G;
 XX WPI: 2002-154726/20.
 XX N-FSDB; ABB90544.
 DR Novel Chlamydia pneumoniae protein useful in the manufacture of a
 PT medicament for treatment or prevention of infection due to Chlamydia,
 PT preferably Chlamydia pneumoniae, and for diagnostic purposes
 PS Claim 5; Page 60-61; 364pp; English.
 XX Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
 CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
 CC them. The proteins are predicted to be immunogenic and may therefore be
 CC useful in vaccine production and for diagnostic purposes. Chlamydia
 CC pneumoniae is a common cause of respiratory disease in humans, and is
 CC also involved in the development of cardiovascular diseases such as
 CC atherosclerosis, coronary artery disease, carotid artery stenosis,
 CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
 CC claudication and stroke. The proteins and nucleic acids of the invention
 CC may be used in vaccines and pharmaceutical compositions for the
 CC prevention or treatment of chlamydial infections, particularly Chlamydia
 CC pneumoniae infections. The proteins may also be used in the detection of
 CC Chlamydia pneumoniae, and the nucleic acids may be used in the detection of
 CC pneumoniae gene expression. The present sequence represents a
 CC specifically claimed DNA which encodes a Chlamydia pneumoniae protein of
 CC the invention.
 XX

SO Sequence 579 BP; 184 A; 131 C; 109 G; 155 T; 0 other;

Alignment Scores:
 Pred. No.: 1,42e-15 Length: 579
 Score: 204.50 Matches: 46
 Percent Similarity: 58.82% Conservative: 24
 Best Local Similarity: 38.66% Mismatches: 46
 Query Match: 23.75% Indels: 3
 DB: 24 Gaps: 2

US-09-674-779B-2 (1-172) x ABL91202 (1-579)

QY 54 AspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSerLeu-----ValTyr 71
 Db 202 GATTCGAAGAGAAAGAAACCAATACAGCAAGCCAGTTCGATATATATCACC 261
 QY 72 PheAspPheAspSerAspGluIleLys---ProGlnAlaAlaIleLeuAspGlu 90
 Db 262 TTTCGTACAGACAGCTATACATTAAGGTAAAGAACCTTGGATTCACGAACCTG 321
 QY 91 AlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGlu 110
 Db 322 GTTCACATCATGAAGAAACCCGAAAGCTACATGATTAAGGAGCATACGACGAG 381
 QY 111 ArgGlySerArgGluTyrAsnMetSerLeuGlyLysArgArgAlaValAlaValArgAsn 130
 Db 382 CGTGAGAGTCATCTCTATACCTTGCTTTAGAGACGACGACGAATGCGATTAAAGAG 441
 QY 131 TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 150
 Db 442 CATCTCGGAAGCAGGGAATCTTCGAGAGTCGTATCTACTATTTCTTCGGAAGAA 501
 QY 151 ArgProIleAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArgArgAlaGlu 169
 Db 502 CATCTTTAAATTCGGGACACACAGCAATGACATGCAACAAATCCCGCTACAGAG 558

RESULT 21
 AAD08593
 ID AAD08593 standard; DNA; 779 BP.
 AC AAD08593;

XX 04-SEP-2001 (first entry)
 DT
 XX Chlamydia pneumoniae omp P6 precursor gene.
 DE
 XX Omp P6 precursor; outer membrane protein; Chlamydia infection;
 KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 KW acute respiratory disease; cough; sore throat; atherosclerosis; asthma;
 KW vaccine; antibiotic; ds.
 XX
 OS Chlamydia pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT CDS 101..679
 FT /*tag= a
 FT /product= "Omp P6 precursor"
 PN MO200146224-A2.
 XX
 PD 28-JUN-2001.
 XX
 PF 20-DEC-2000; 2000WO-CA01534.
 XX
 PR 22-DEC-1999; 99US-0171525.
 XX
 XX (AVET) AVENTIS PASTEUR LTD.
 XX
 PL Murdin AD, Oomen RP, Wang J, Dunn P;
 XX
 DR WPI; 2001-408630/43.
 XX
 DR P-PSDB; AAE04322.
 XX
 PT Chlamydia polypeptides and polynucleotides for preventing, diagnosing
 XX and treating Chlamydia infection in mammals, especially humans -
 XX
 PS Claim 2; Fig 1; 74pp; English.
 XX
 CC The present sequence is Chlamydia pneumoniae outer membrane
 CC protein (omp) P6 precursor gene. The omp P6 precursor polynucleotide,
 CC polypeptide and its antibody are useful for detecting, preventing and
 CC treating Chlamydia infections e.g. pneumonia, upper respiratory tract
 CC diseases (e.g. bronchitis, sinusitis) and acute respiratory diseases
 CC (e.g. cough, sore throat). C. pneumoniae infection is also observed to
 CC be associated with atherosclerosis and asthma. Immunogenic fragments
 CC of omp P6 precursor protein and their corresponding DNAs are useful in
 CC vaccine composition. The polynucleotide is useful in the construction
 CC of attenuated Chlamydia strains that can over express the polynucleotide
 CC or express it in a non-toxic, mutated form. The omp P6 probes are useful
 CC in diagnostic tests as capture or detection probes and the primers are
 CC useful in diagnostic methods involving PCR (polymerase chain reaction).
 CC The antibody is useful for purifying polypeptide.
 XX
 SQ Sequence 779 BP; 240 A; 175 C; 150 G; 214 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 2,14e-15 Length: 779
 Score: 204.50 Matches: 46
 Percent Similarity: 58.82% Conservative: 24
 Best Local Similarity: 38.66% Mismatches: 46
 Query Match: 23.75% Indels: 3
 DB: 22 Gaps: 2
 US-09-674-779B-2 (1-172) x AAD08593 (1-779)
 QY 54 ASPASNAAPGJUTHTVALVYALALEUASERLYSLEUPROSERLEU-----VALTYR 71
 DB 302 GATTCACAAAGAGAAAAACATACAGTCAAGCCAAAGTTCGATTTCCGAATATACACC 361
 QY 72 PheaspPheaspSeraspGluIleLys---ProGlnAlaAlaAlaIleLeuaspGluGln 90
 DB 362 TTTCGTACAGACAGCTATACATTAAGGTGAGAGAACCTTCGATTCACGAACTTG 421
 QY 91 AlagInPheLeuThThAsnGlnThThAlaArgValLeuValAlaIagIyHsThAspGlu 110

DB 422 GTTCACATCATGAGAAAAACCCGAAGCTACACTGTACATTGAGAGCATACTGAGAG 481
 QY 111 ArgGlySerArgGlyUTyTAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsn 130
 DB 482 CGTGGAGCTGCATCTCTATACCTTGCTTACGACACGACGACCAATGGATTAAGAG 541
 QY 131 TyTLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 150
 DB 542 CATCTCCGAAGACGAGGAATCTGTGACAGATGCTATCTACTACTTCTACGGAAAAAGAA 601
 QY 151 ArgProIleAlaPheGlyThAsnGluGluAlaTrpSerGlnAsnArgAlaGlu 169
 DB 602 CATCCTTTAAATTCGCGACACAGCACTGACATGCAACAAATGCCGTACAGAG 658
 RESULT 22
 ID AAX91990/c
 ID AAX91990 standard; DNA; 1230025 BP.
 AC AAX91990;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; ss.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN WO9227105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 XX
 PR 21-NOV-1997; 97FR-0014673.
 XX
 XX (GEST) GENSET.
 XX
 PA Griffais R;
 XX
 DR WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 XX
 PS Claim 1; Page 291-611; 1912pp; English.
 XX
 CC The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
 XX
 Alignment Scores:
 Pred. No.: 6.07e-11 Length: 1230025
 Score: 204.50 Matches: 46
 Percent Similarity: 58.82% Conservative: 24
 Best Local Similarity: 38.66% Mismatches: 46
 Query Match: 23.75% Indels: 3
 DB: 20 Gaps: 2
 US-09-674-779B-2 (1-172) x AAX91990 (1-1230025)

[illegible]

Db 3593 CAGAGCTGTCGACGCGTGTGCGCCACGCGCTGTGACGACTGCTACAGACTACGCT 3652
 QY 137 ILeaNgInAlaSerValGluIleIleSerPheGlyLyuGuaArgProIleAlaPheGly 156
 Db 3653 GTAAGAAGTGTGCTCCGTAACGCTGTGTACGCGAGTCCGCCGCGTTGCCGACMAC 3712
 QY 157 ThrAsnGluAlaTrpSerGlnAsnArgArgAlaGlu 169
 Db 3713 GCCACGCGTGAAGCCGCGCTATCAACGCTGCGTTGAA 3751

RESULT 29

AAT32599

ID AAT32599 standard; cDNA; 486 BP.

AC AAT32599;

DT 01-OCT-1996 (first entry)

DE P. aeruginosa OprF C-terminal region cDNA.

KW Vaccine; fusion protein; OprF; OprI; outer membrane protein;

XX antibody; glutathione-S-transferase; GST; ds.

OS Pseudomonas aeruginosa serotype 6 (ATCC 33354).

PN EP717106-A1.

PD 19-JUN-1996.

XX 17-NOV-1995; 95EP-0118098.

PR 16-DEC-1994; 94EP-0120023.

PA (BEHW) BEHRINGWERKE AG.

PI Broecker M, Domdey H, Hungerer K, Knapp B, Ulrich B;

DR WPI: 1996-279559/29.

XX P-PSDB; AAR99625.

PT Fusion proteins for Pseudomonas aeruginosa vaccines - contg.

PS Fragments of outer membrane proteins I and F.

XX Disclosure; Page 11-12; 23pp; English.

CC A cDNA sequence (AAT32599) codes for amino acids 190-350 (AAR99625) of

CC the Pseudomonas aeruginosa ATCC 33354 outer membrane protein F

CC (OprF), i.e. the C-terminal region of the protein. It was obt.

CC by PCR amplification of the OprF gene in plasmid pFsaui (see also

CC AAT32592-93). A hybrid gene between glutathione-S-transferase and

CC the OprF sequence was expressed in *Escherichia coli*. OprF-OprI

CC (AAT32600) and OprI-OprF (AAT32600) hybrid genes were constructed and

CC the encoded proteins (AAR99626-27) were tested for their efficacy as

CC vaccines against *P. aeruginosa* infection.

XX Sequence 486 BP; 109 A; 158 C; 138 G; 81 T; 0 other;

Alignment Scores:

Pred. No.: 1,2e-11 Length: 486
 Score: 171.50 Matches: 48
 Percent Similarity: 46.15% Conservative: 18
 Best Local Similarity: 33.57% Mismatches: 67
 Query Match: 19.92% Indels: 11
 DB: 17 Gaps: 2

US-09-674-779B-2 (1-172) x AAT32599 (1-486)

QY 37 AlaProThrgLyTrpThrgLyValIleTyTrhGlyValAlaProLeuValAspAsnAsp 56
 Db 32 GTCGCGACTCCGACGACGCGGCTGTGGGACAC-GTCGACAAGTCCCGACACCCCG 90

QY 57 GluThrValIleAlaLeuAlaSerLysLeuProSerLeu----- 69
 Db 91 GCCACGCTACCGCTGACCCCAACGCGCTGCCGCTGCGGAAGCTGTACGCTACAG 150
 QY 70 -----ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeu 87
 Db 151 CTGAGAGTGAAGTGCACCTTCGACCAAGTCCAAAGAGACACAGCTACGCTACATC 210
 QY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHis 107
 Db 211 AAGAACCTGCGGACCTCATGAAGCAGTACCGCTCCACTTCACACCGCTTGAAGTCAT 270
 QY 108 ThrAspGluArgGlySerArgGlyLysArgLysAsnMetSerLeuGlyGluArgAlaValAla 127
 Db 271 ACCGACTCCGCTGCTGACGACGCTTACCAACGAAAGCTGCGAGCTGCCAACGCC 330
 QY 128 ValArgAsnTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleIleSer 146
 Db 331 GTTCGTGAGCTAGCTGTGTCACAGTACGAGTGTGGAAGTGTGCGGTACACGCTGCGGT 390
 QY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166
 Db 391 TACGGGAGTCCCGCGCGGTGCCGACCAACGCCACCGCTGAAGGCCGCTATCAACGCT 450
 QY 167 ArgAlaGlu 169
 Db 451 CGCGTTGAA 459

RESULT 30

AAT34419

ID AAT34419 standard; cDNA; 681 BP.

AC AAT34419;

DT 01-OCT-1996 (first entry)

DE P. aeruginosa OprI-OprF hybrid gene.

KW Vaccine; fusion protein; OprF; OprI; outer membrane protein;

XX antibody; glutathione-S-transferase; GST; ds.

OS Pseudomonas aeruginosa serotype 6 (ATCC 33354).

PN EP717106-A1.

PD 19-JUN-1996.

XX 17-NOV-1995; 95EP-0118098.

PR 16-DEC-1994; 94EP-0120023.

PA (BEHW) BEHRINGWERKE AG.

PI Broecker M, Domdey H, Hungerer K, Knapp B, Ulrich B;

DR WPI: 1996-279559/29.

XX P-PSDB; AAR99627.

PT Fusion proteins for Pseudomonas aeruginosa vaccines - contg.

PS Fragments of outer membrane proteins I and F.

XX Disclosure; Page 15-16; 23pp; English.

A cDNA sequence (AAT34419) codes for a fusion (AAR99627) between amino
 acids 21-83 of the outer membrane protein I (OprI) and amino acids
 190-350 of outer membrane protein F (OprF) of *Pseudomonas aeruginosa*
 ATCC 33354, the 2 moieties being separated by a dipeptide linker. A
 CC glutathione-S-transferase fusion with the hybrid gene was expressed
 CC in *Escherichia coli* and *Saccharomyces cerevisiae* transformants. The
 CC oprI-oprF hybrid protein was significantly less immunogenic than an
 CC oprF-oprI fusion (AAR99626).

Sequence 681 BP; 158 A; 221 C; 195 G; 107 T; 0 other;

Alignment Scores:

Pred. No.:	1,91e-11	Length:	681
Score:	171.50	Matches:	48
Percent Similarity:	46.15%	Conservative:	18
Best Local Similarity:	33.57%	Mismatches:	67
Query Match:	19.92%	Indels:	11
	17	Gaps:	2

US-09-674-779b-2 (1-172) x AAN34419 (1-681)

```

QY 37 AlaProthrglytyrthrglyvalilleyrthrglyvalalaproleuvalaspasnasp 56
    |||||
Db 227 GCTCCGACCTCCGACACGACGAGGGGTCTGCCACAC-ctgcacaaagtcggccgacaccccg 285
    |||||
QY 57 GluThrVallysalaleualaserlysluProserleu----- 69
    |||||
Db 286 GCCAACGTCACCGCTTGACGCCAACGGCTGCCGCTGCCGCAAGCTCGTACCGCTACAG 345
    |||||
QY 70 -----ValtyrPheaspPheaspserAspGluilleysProGlnAlaAlaIleleu 87
    |||||
Db 346 CTGACGCTGAAGTTCGCTCGCTCGACAGTCCAGAGTCAAGAGACAGCTACGCTGACATC 405
    |||||
QY 88 AspGluGlnAlaGlnPheleuthrthrasnglnThrAlaArgValleuValAlaGlyHis 107
    |||||
Db 406 AAGAACCTGGCGGCTCATGATGAAGACAGTACCCGCTCCACCTCCACCGCTGAAGTCTAT 465
    |||||
QY 108 ThrAspGluArglySerArgGlyTyrAsnmetSerleuGlyGluArgArgAlaValAla 127
    |||||
Db 466 ACCGACTCCGCTCGCTGACGACGCTTACACCAAGCTGTCGACGCTCGTCCACAGCC 525
    |||||
QY 128 ValArgAsnTyrleuGlyLys---GlyTlleasnglnAlaserValGluilleSer 146
    |||||
Db 526 GTTCGTACAGCTACGCTGACAGAGTACGAGTGTGGAAGCTGCTGCGTGAACGCTGTCGG 585
    |||||
QY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166
    |||||
Db 586 TACGGGAGTCCCGCGGTGTCGACCAAGCCACCGCTGAAGGCGCGCTATCAACCGT 645
    |||||
QY 167 ArgAlaGlu 169
    |||||
Db 646 CGCGTTGAA 654
    |||||

RESULT 31
AAN82023
ID AAN82023 standard; DNA; 1253 BP.
XX
AC AAN82023:
XX
DT 22-OCT-1990 (first entry)
XX
DE Outer membrane protein F of Pseudomonas aeruginosa.
XX
KW Outer membrane protein F; OMPF; vaccination; antibodies; ss.
XX
OS Pseudomonas aeruginosa.
XX
FH Key location/Qualifiers
FT CDS 64..1114
    /tag=a
    /product=Outer membrane protein F
XX
PD DE3718591-A.
XX
PF 15-DEC-1987.
XX
PR 03-JUN-1987; 87DE-3718591.
XX
PA (BEHW ) BEHRINGWERKE AG.
XX

```

PI Domdey H, Lottspeich F, von Specht B-U, Duchene M;
 XX WPI: 1988-361619/51.
 DR P-PSDB; AAP82053.

XX New outer membrane protein F of Pseudomonas aeruginosa -
 PT DNA sequences encoding it and derived antibodies, useful for
 PR vaccination and diagnosis.
 PS Disclosure; p; German.

XX The protein is isolated from the OMP of P. aeruginosa serotype 6
 CC ATTC 33354) and purified by HPLC. The amino-terminal and trypsin
 CC fragments are sequenced and a series of oligonucleotide probes
 CC constructed corresponding to the established sequences. These probes are
 CC used to screen a gene bank of 15-20 kb fragments of genomic DNA in
 CC lambda Emu 3. One positive clone includes a 15 kb insert contg. the
 CC protein gene, which can be isolated as a 2.5 kb peti fragment. This
 CC fragment cannot be cloned into a high copy no. vector because of the
 CC toxicity of the gene prod., so is subcloned as two fragments with an
 CC overlapping region of about 500bp. Ab's are raised by usual immunisation
 CC or cell-fusion procedures. The DNA is useful in diagnosis.

Sequence 1253 BP; 279 A; 389 C; 358 G; 227 T; 0 other;

Alignment Scores:

Pred. No.:	4.47e-11	Length:	1253
Score:	171.50	Matches:	56
Percent Similarity:	45.36%	Conservative:	27
Best Local Similarity:	30.60%	Mismatches:	79
Query Match:	19.92%	Indels:	22
	9	Gaps:	4

US-09-674-779b-2 (1-172) x AAN82023 (1-1253)

```

QY 8 AlaAlaAlaAlaAlaAlaLeuSerValleuthrPheMetThrGly-----CysAla 24
    |||||
Db 542 ACTACGAGCTTGAGAGAGCTGTGACACAGCTCACAGGATGATGGCTGCTGGCGGCG 601
    |||||
QY 25 AsnLysSerThrSerGlnVal-----MetValAlaProAsn 36
    |||||
Db 602 TCGGCTTCACACTTGGTGTGGAAGCCGCTCCGCTCCGGAACCGGTGTCGACAGTTT 661
    |||||
QY 37 AlaProthrglytyrthrglyvalilleyrthrglyvalalaproleuvalaspasnasp 56
    |||||
Db 662 GCTCCGACTCCGACACAGCGCGCTCTGCGACAC-ctgcacaaagtcggccgacaccccg 720
    |||||
QY 57 GluThrVallysalaleualaserlysluProserleu----- 69
    |||||
Db 721 GCCAACGTCACCGCTTGACGCCAACGGCTGCCGCTGCCGCAAGCTCGTACCGGTACAG 780
    |||||
QY 70 -----ValtyrPheaspPheaspserAspGluilleysProGlnAlaAlaIleleu 87
    |||||
Db 781 CTGACGCTGAAGTTCGACTTCGACAAAGTTCACAAAGAGAACGCTACGCTGACATC 840
    |||||
QY 88 AspGluGlnAlaGlnPheleuthrthrasnglnThrAlaArgValleuValAlaGlyHis 107
    |||||
Db 841 AAGAACCTGGCGGCTCATGATGAAGACAGTACCCGCTCCACTCCACCGCTGAAGCTAT 900
    |||||
QY 108 ThrAspGluArglySerArgGlyTyrAsnmetSerleuGlyGluArgArgAlaValAla 127
    |||||
Db 901 ACCGACTCCGCTGACACGACGCTTACACCAAGAGTGTGTCGACGCTGTCGCAACCC 960
    |||||
QY 128 ValArgAsnTyrleuGlyLys---GlyTlleasnglnAlaserValGluilleSer 146
    |||||
Db 961 GTTCGTACAGCTACGCTGACAGAGTGTGGAAGTGTGCGTGAACGCTGTCGCT 1020
    |||||
QY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166
    |||||
Db 1021 TACGGGAGTCCCGCGGTGTCGACCAAGCCACCGCTGAAGGCGCGCTATCAACCGT 1080
    |||||
QY 167 ArgAlaGlu 169
    |||||

```


DB 1081 CGCGTTGAA 1089

RESULT 32

AAA38558

ID AAA38558 standard; DNA; 1319 BP.

XX AAA38558;

AC

XX 11-SEP-2000 (first entry)

DT

XX

DE Actinobacillus pleuropneumoniae ompA2 DNA.

XX

KW OmpA2; outer membrane protein; APP; pneumonic pathogen; swine;

KW Sus scrofa; serotype; antigen; vaccine; cross-reactive immunity;

KW cross-protection; diagnosis; ds.

XX

OS Actinobacillus pleuropneumoniae strain Pz420.

XX

PH Key Location/Qualifiers

FT CDS . 197..1306

FT /tag= a

FT /product= "Actinobacillus pleuropneumoniae (APP) OmpA2"

FT sig_peptide 297..253

FT /tag= b

FT mat_peptide 254..1303

FT /tag= c

XX

PN EP1001025-A2.

PD 17-MAY-2000.

XX

PF 20-OCT-1999; 99EP-0308262.

XX

PR 22-OCT-1998; 98US-0105285.

XX

PA (PF12) PFIZER PROD INC.

XX

PI Ankenbauer RG, Baarsch MJ, Campos M, Kelch RL, Rosey EL;

PI Warren-Stewart LM, Sulter BT;

XX

DR WPI: 2000-320438/28.

DR P-PSDB: AAY97900.

XX

PT Low molecular weight Actinobacillus pleuropneumoniae proteins and DNA

PT encoding them, for use as vaccines against the bacteria in swine -

XX

PS Claim 32; Page 49-51; 81pp; English.

XX

CC This sequence represents DNA encoding a low molecular weight outer

CC membrane protein, ompA2, from Actinobacillus pleuropneumoniae (APP)

CC strain Pz420 (ATCC 98930). The invention relates to the novel APP outer

CC membrane proteins omp20, omp27, ompA1 and ompA2 (AAY97896-Y97900)

CC and to nucleic acids encoding them (AAA38554-A38558). APP is a Gram

CC negative coccobacillus which is one of the most important swine

CC pneumonic pathogens. 12 different serotypes of APP have been recognised

CC which vary in geographic distribution. Prior art attempts at vaccinating

CC against APP have produced mainly serotype-specific immune responses. In

CC contrast, natural immunity to any one serotype seems to confer

CC significant protection from disease caused by other serotypes, suggesting

CC that natural exposure induces cross-reactive immunity to shared antigens.

CC The novel outer membrane proteins of the invention are present in all 12

CC serotypes, and may provide a target for cross-protective immunisation.

CC The novel outer membrane proteins and nucleic acids encoding them can be

CC used as a vaccine against APP in swine. They can also be used as

CC reagents for the diagnosis of APP infections.

XX

SO Sequence 1319 BP; 428 A; 252 C; 260 G; 379 T; 0 other;

Alignment Scores:

Pred. No.: 7.32e-11

Score: 170.00

Percent Similarity: 48.59%

Best Local Similarity: 34.51%

Length: 1319

Matches: 49

Conservative: 20

Mismatches: 61

Query Match: 19.74% Indels: 12

DB: 21 Gaps: 5

US-09-674-779b-2 (1-172) x AAA38558 (1-1319)

QY 34 AAlapRoasnaAlaProthrgLYTYrThrGlyValIleTYrThr-----GlyVal 49

DB 824 GCTCCTGATATCCACTGCTGTAACAGCAGGTTTATCATATCGTTCCGTCAGCGCTGTA 883

QY 50 AAlapRoLeuValAspasnaspgLutThrValIleValAlaLeuAlaSerLysLeuProSerLeu 69

DB 884 GCACGAGTTGTTGAGCCAGAGTGTAACTAAAACTTCGCA-----TTCAGCTCAGAC 937

QY 70 ValIyrPheaspPheaspSerAspLuiIleYsrProGlnAlaAlaIleLeuAspGlu 89

DB 938 GTTTTATTGTTGTTGGTAAATCAAGCTTAAACACGACGACGACACAGCTTATGAC-- 994

QY 90 GlnAlaGlnPheLeuThrThrAsn-----GlnThrAlaArgValIleValAlaGly 106

DB 995 GCAGCTAACTAGTAAGAAACGCTAACTAGTTAGCTTACCACTCCAGCTATCCAACTTAACGT 1054

QY 107 HisThrAspGluArgGlySerArgGluTrpAsnMetSerLeuGlyGluArgAlaVal 126

DB 1055 TATACAGACCGCTATCGGTAAAGAAAGCTTCAAACTTAACTTCAACACCGCTGCAGAA 1114

QY 127 AlaValArgAsnTrpLeuLeuGlyLysGlyIleAsnGlnIleAsrValGluIleSer 146

DB 1115 ACTGTAGCTACTACTTATGTTTCTTAAGAGTCAAAACCTGCAAACTGCACTGCAATGCT 1174

QY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166

DB 1175 TAGGTGAGCAAAACCCAGTACCGGGCGCAACATGTACAAAGTT-----AAAGTCTGT 1228

QY 167 ArgAla 168

DB 1229 AAAGCA 1234

RESULT 33

AACT9664

ID AACT9664 standard; DNA; 1110 BP.

XX

AC AACT9664;

XX

DT 08-FEB-2001 (first entry)

XX

DE Virulence gene #71.

XX

KW Virulence gene; antibacterial; vaccine; bacterial infection;

KW septicaemia; bronchopneumonia; rhinitis; wound infection; ss.

XX

OS Actinobacillus pleuropneumoniae.

XX

PN WO200061724-A2.

XX

PD 19-OCT-2000.

XX

PE 06-APR-2000; 2000WO-US09218.

XX

PR 09-APR-1999; 99US-0128689.

PR 10-SEP-1999; 99US-0153453.

XX

PA (PHAA) PHARMACIA & UPJOHN INC.

XX

PI Lowery DE, Fuller TE, Kennedy MJ;

XX

DR WPI: 2000-647422/62.

DR P-PSDB: AAB44589.

XX

PT Attenuated Pasteurellaceae bacteria comprising mutations in virulence

PT genes, useful as a live attenuated vaccine against bacterial infections

XX

PS Claim 7; Pages 306-308; 322pp; English.

XX The family Pasteurellaceae encompasses several pathogens that infect a
 CC wide variety of animals. The present invention relates to virulence genes
 CC from Pasteurellaceae. The present sequence is one such virulence gene.
 CC The present sequence may be mutated in order to produce an inactive gene.
 CC The inactive virulence gene may in turn be used to produce a vaccine,
 CC which is useful for treating bacterial infections such as septicemias,
 CC bronchopneumonias, rhinitis and wound infections.

XX
 XX Sequence 1110 BP; 365 A; 229 C; 220 G; 296 T; 0 other;

Alignment Scores:
 Pred. No.: 7.63e-11 Length: 1110
 Score: 169.00 Matches: 47
 Percent Similarity: 48.09% Conservative: 16
 Best Local Similarity: 35.88% Mismatches: 58
 Query Match: 19.63% Indels: 10
 DB: 21 Gaps: 4

US-09-674-779b-2 (1-172) x AAC79664 (1-1110)

QY 34 AlapProasnaIapProthGlyTyrThrValIleTyrThr-----GlyVal 49
 |||||
 DB 628 GCTCCTGATATCCACTCTGTACAGCAGGTTTATCATACCGTTTCGCTCAAGCGCTGTA 687
 QY 50 AlapProleuValaspasnaSpGluThrValIysAlaLeuAlaSerLysLeuProSerLeu 69
 |||||
 DB 668 GCACACAGTGTGTGAGCCAGAGTGTACTATAAAACCTGCA-----TTCAGCTCAGAC 741
 QY 70 ValIyrPheaspPheaspSeraspGluIleLysProGlnAlaAlaIleLeuaspGlu 89
 |||||
 DB 742 GTTATTATGATTGTTGCGTAATCAAGCTTAACACACAGCAACGCTTAGAC--- 798
 QY 90 GlnIaGlnPheLeuThrThrAsn-----GlnThrAlaArgValLeuValAlaGly 106
 |||||
 DB 799 GCAGCTAACACTGAATTCGCTAACTTACGTTTACCACTCCAGCTATCCAACTTAACGGT 858
 QY 107 HisThrAspGluArgIysSerArgIuTyraSmetSerLeuGlyGluArgAlaVal 126
 |||||
 DB 859 TATACAGACCGTAAAGGTAAGAGCTTAACCTTAACCTTCAACAGCCGCGAGAA 918
 QY 127 AlaValaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSer 146
 |||||
 DB 919 ACTGTAAGCTAACTTAAGTTCCTTAAGGTCAAAACCCGCAACGTAACGTGACAGTAGGT 978
 QY 147 PheGlyGluGluArgProIleAlaPheGlyThr 157
 |||||
 DB 979 TACGCTGAAGCAACCCAGTAAACCGCGCAACA 1011

RESULT 34
 AAQ78916
 ID AAQ78916 standard; DNA; 1720 BP.
 AC AAQ78916;
 XX
 DT 09-AUG-1995 (first entry)
 XX
 DE Non-typable Haemophilus influenza (NTH1) fimbria gene.
 XX
 KW Fimbria protein; vaccine; otitis media; ss.
 XX
 OS Haemophilus influenza strain 1128.
 XX
 Key Location/Qualifiers
 FH 406..1485
 FT CDS /tag= a
 FT RBS 385..388
 FT /tag= b
 FT stem_loop 1518..1547
 FT /tag= c
 XX
 PN MO9426304-A.

PD 24-NOV-1994.
 XX
 PF 12-MAY-1994; 94MO-US05477.
 XX
 PR 18-MAY-1993; 93US-0065442.
 XX
 PA (OHIO-) OHIO STATE RES FOUND.
 XX
 PI Bakaletz LO, Kolattukudy PE, Strakova T;
 XX
 DR WPI: 1995-006359/01.
 DR P-PSDB; AAR66294.
 XX
 PT Vaccine comprising non-typable Haemophilus influenza fimbria
 PT protein - useful in studying, preventing or reducing the
 PT severity of otitis media, also fimbria protein and DNA.
 PS Claim 12; Fig 5; 45pp; English.

XX
 XX The fimbria proteins from 15 randomly selected type b and non-
 CC typable clinical isolates of Haemophilus influenza share common
 CC epitopes. Thus fimbria isolated from non-typable Haemophilus
 CC influenzae 1128 strain is a particularly suitable immunogen to
 CC protect against the different non-typable H₂ influenzae that cause
 CC otitis media. Fimbria protein is produced by culturing a transformed
 CC microbial host, pref. E. coli, sporodoptera frugiperda or a mucosal
 CC pathogen. Fimbria protein (FP) produced by this process is claimed.
 CC The FP protein migrates in polyacrylamide gels to a posn. equiv. to
 CC a mol. wt. of 25.5 kd or 37.5 kd.

XX
 SQ Sequence 1720 BP; 571 A; 301 C; 343 G; 505 T; 0 other;

Alignment Scores:
 Pred. No.: 3.76e-10 Length: 1720
 Score: 165.50 Matches: 47
 Percent Similarity: 46.50% Conservative: 26
 Best Local Similarity: 29.94% Mismatches: 63
 Query Match: 19.22% Indels: 21
 DB: 16 Gaps: 4

US-09-674-779b-2 (1-172) x AAQ78916 (1-1720)

QY 35 ProasnaIapProthGlyTyr-----ThrGlyValIleTyr 46
 |||||
 DB 1000 CCAATATCCGCAATTAATCAACCTTGATGTTGATCAATGCGGATATTTCTTAC 1059
 QY 47 -----ThrGlyValaIapProleuValaspasnaSpGluThrValIysAlaLeuAla 63
 |||||
 DB 1060 CGTTTGGTCAAGGCGACAGCAGCTGTGTGACAGCACTGAATGTAAGCAAACTTTC 1119
 QY 64 SerLysLeuProSerLeuValTyrPheaspPheaspSeraspGluIleLysProGlnAla 83
 |||||
 DB 1120 AGC---TTAAATTCGTATGTAATCTTCCGATTTGTTAAAGCAAACTTCAACCA 1176
 QY 84 AlaAlaIleLeuaspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeu 103
 |||||
 DB 1177 CAAGCTACATTAAGTACGCGCTATGCGCAAAATTTCACAAGTTAAAGTCGAAGTACCT 1236
 QY 104 ValaIaGlyHisThrAspGluArgIysSerArgIuTyraSmetSerLeuGlyGluArg 123
 |||||
 DB 1237 GTTGGCTGTACACTAACCGTATGTTGTTGACGCGGTTCACAGTAACTTCTCAACA 1296
 QY 124 ArgAlaValaIaValaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGlu 143
 |||||
 DB 1297 CGTGCAGATTCAGTACCTAACTTGTGCTAAAGGTGTGACAGCAGCAATCTCA 1356
 QY 144 IleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu----- 160
 |||||
 DB 1357 GCAAACTGTTACGGTGAACCAACCCAGTAACTGCGGCAACTGTGACCAAGTTAAAGCT 1416
 QY 161 -----AlaTrpSerGlnAsnArgArgAlaGluLeuSer 171
 |||||
 DB 1417 CGTAAAGCACTTATGCTTGTCTGCTCCAGACCGTGTGTAGAAATCGCA 1467

RESULT 35
AAI46503
ID AAI46503 standard; DNA; 898 BP.
XX
AC AAI46503;
XX
DT 19-JUL-2002 (first entry)
XX
M catarrhalis MCA100854 gene SEQ ID NO: 21.
XX
KW Moraxella; vaccine; respiratory tract infection; antiinflammatory; auditory; antibacterial; otitis media; sinusitis; pneumonia; gene; ds.
OS Moraxella catarrhalis.
XX
PX MO200218595-A2.
XX
PD 07-MAR-2002.
XX
PF 28-AUG-2001; 2001MO-CA01221.
XX
PR 28-AUG-2000; 2000US-228294P.
PR 28-AUG-2000; 2000US-228295P.
PR 28-AUG-2000; 2000US-228296P.
PR 29-AUG-2000; 2000US-228438P.
PR 29-AUG-2000; 2000US-228439P.
PR 29-AUG-2000; 2000US-228440P.
PR 29-AUG-2000; 2000US-228441P.
PR 29-AUG-2000; 2000US-228442P.
PR 29-AUG-2000; 2000US-228443P.
PR 29-AUG-2000; 2000US-228511P.
PR 29-AUG-2000; 2000US-228512P.
PR 29-AUG-2000; 2000US-228743P.
PR 29-AUG-2000; 2000US-228773P.
PR 01-SEP-2000; 2000US-229465P.
PR 01-SEP-2000; 2000US-229474P.
PR 01-SEP-2000; 2000US-229475P.
PR 01-SEP-2000; 2000US-229478P.
PR 05-SEP-2000; 2000US-229740P.
PR 05-SEP-2000; 2000US-229803P.
PR 05-SEP-2000; 2000US-229804P.
PR 05-SEP-2000; 2000US-229805P.
PR 05-SEP-2000; 2000US-229806P.
PR 05-SEP-2000; 2000US-229809P.
PR 05-SEP-2000; 2000US-229811P.
PR 06-SEP-2000; 2000US-230214P.
PR 06-SEP-2000; 2000US-230250P.
PR 06-SEP-2000; 2000US-230252P.
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;
XX
DR WPI: 2002-401721/43.
DR P-PSDB; MAOI1571.
XX
PT Moraxella polypeptide and polynucleotides useful as vaccine for immunizing a host e.g. humans against disease e.g. otitis media, pneumonia, caused by infection of the bacteria -
PT
XX
PS Claim 2; Fig 21; 277pp; English.
XX
The present invention provides the protein and coding sequences of CC proteins from Moraxella catarrhalis. These can be used to produce CC vaccines which protect against M. catarrhalis infection, which can CC cause otitis media, respiratory infection, sinusitis, and pneumonia. The CC present sequence is a coding sequence of the invention. XX
SQ Sequence 898 BP; 290 A; 171 C; 182 G; 255 T; 0 other;
Alignment Scores: 1.75e-10 Length: 898 Pred. No.:

Score:	165.00	Matches:	42
Percent Similarity:	46.72%	Conservative:	22
Best Local Similarity:	30.66%	Mismatches:	71
Query Match:	19.16%	Indels:	2
DB:	24	Gaps:	1
US-09-674-7798-2 (1-172) x AAL46503 (1-898)			
QY	33	ValAlaProAsnAlaProthrhnglylYrthrhnglyValIleYrthrhnglyValAlaProLeu	52
DB	302	GTTCGCGATGATGTTGAAACACGCGGTGGTATGGTATGGCTTTGGT-----GATACG	355
QY	53	ValAspAsnAspGluThrVallylsAlaLeuAlaSerLysLeuProSerLeuValTyrPhe	72
DB	356	GTTGACAGAAATAGATGGATATTTTAAATCAATGCATTAAATACCCAAATCATTAACCTT	415
QY	73	AspPheAspSerAspGluIleYsProGlnAlaAlaAlaIleLeuAspGluGln	92
DB	416	GCTTTAAACATCAACCCAAATTCGCCAGMAAAATTAAGAAATCTTGATTTGGCCGAA	475
QY	93	PheLeuThrthrasnglnThrAlaArgValLeuAlaGlyYHsthrAspGluArgGly	112
DB	476	AAATTTAAGGCACTGCCCTGAACACTTTCGATATCGATATGGCATACAGACACTCAAGGC	535
QY	113	SerArgIuTyrAsnMetSerLeuGluArgArgAlaValAlaValArgAsnTyrLeu	132
DB	536	ACCGATGAGTATATCAAGATTTATCAGAAATCTCGTCTGCTGCTGTTAAAGACTATTTG	595
QY	133	LeuGlyIysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgPro	152
DB	596	GATCAACAAAGTGTCTCTGCTGACGCTTGAACACTCAAGGTGCAAGTTTGATTATCCA	655
QY	153	IleAlaPheGlyThrAsnGluGluAlaTyrSerGlnAsnArgArgAlaGlu	169
DB	656	GTTGCATCAAAATGCTACGACAGACAGGTGCTTCAAAACCGTCGATTGGAG	706
RESULT 36			
AAAF28548			
ID	AAAF28548 standard; DNA: 96109 BP.		
XX	AAAF28548;		
AC	AAAF28548;		
XX	04-APR-2001 (first entry)		
DT	04-APR-2001 (first entry)		
XX	Genomic fragment #35.		
DE	Genomic library; bacteria; human upper airway; otitis media; sinusitis;		
XX	Genomic library; bacteria; human upper airway; otitis media; sinusitis;		
KW	bronchopulmonary; endocarditis; meningitis; ss.		
XX	Moraxella catarrhalis.		
OS	Moraxella catarrhalis.		
XX	W0200078968-A2.		
PN	W0200078968-A2.		
XX	28-DEC-2000.		
PD	28-DEC-2000.		
XX	16-JUN-2000; 2000MO-US16649.		
PF	16-JUN-2000; 2000MO-US16649.		
XX	18-JUN-1999; 99US-0140121.		
PR	18-JUN-1999; 99US-0140121.		
XX	(INCYTE GENOMICS INC.		
XX	(INCYTE GENOMICS INC.		
PA	(INCYTE GENOMICS INC.		
XX	Lagace RE, Patterson C, Berg KL;		
PI	Lagace RE, Patterson C, Berg KL;		
XX	WPI; 2001-041427/05.		
DR	WPI; 2001-041427/05.		
XX	Genomic library for identifying diagnostic and therapeutic		
PT	compositions, and for identifying virulence factors, regulatory		
PT	elements and drug targets, comprises Moraxella catarrhalis nucleic		
PT	acids -		
XX	Claim 1; Page 345-368; 545pp; English.		
ES	Claim 1; Page 345-368; 545pp; English.		
XX	The present invention relates to a Moraxella catarrhalis genomic library		
CC	The present invention relates to a Moraxella catarrhalis genomic library		

comprising of a combination of 41 nucleic acid molecules (see AAF28514-AAF28554). The library has a number of uses described in the specification e.g. is useful for identifying diagnostic and therapeutic compositions. *M. catarrhalis* (Branhamella catarrhalis) is a large aerobic, gram-negative diplococcus, normally found among the bacterial flora of human upper airways. *M. catarrhalis* is known to cause acute, localised infections such as otitis media, sinusitis and bronchopulmonary infection and life-threatening, systemic diseases including endocarditis and meningitis.

Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1.17e-07	96109	165/00	42	22	71	2
Score:		46.72%		30.66%		19.16%
Best Local Similarity:						
Query Match:						

US-09-674-779B-2 (1-172) x AAF28548 (1-96109)

```

OY 33 ValAlaProAsnAlaProThrGlyTyrThGlyValIleTyrThrGlyValAlaProLeu 52
    |||||  ::::|  |||  ::::|  |||
Db 68342 GTTCCCGATGCTATTGAAACGCCGCTGCTATTGCTGCTTTGGT-----GATACG 68395
OY 53 ValAspAsnAspGluThrValIleValAlaLeuAlaSerIleuProSerLeuValTyrPhe 72
    |||||  ::::|  ::::|  ::::|  ::::|
Db 68386 GTTAGAAGAAATGAGATGATTTTATCAATGATTAATACCCAAATCATTAACCTT 68455
OY 73 AspPheAspSerAspGluIleValLeuValProGlnAlaIleLeuAspGluGlnAlaGln 92
    |||||  |||||  ::::|  ::::|  |||||  |||||
Db 68456 GCTTTAGACCTCAACCCAAATTTCCCAAGAAATTAAGAAATCTGATTTGGCTGCCGAA 68515
OY 93 PheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGly 112
    |||  |||  ::::|  ::::|  ::::|  ::::|  ::::|
Db 68516 AATTTAAAGGACGCTCCGTAAGCAACTTTCGCTGTCATTTGCTCTACGACACTCAAGGC 68575
OY 113 SerArgGluTyrAsnMetSerLeuGlyGluArgGlnAlaValAlaValArgAsnTyrLeu 132
    ::::|  |||||  |||  |||  |||||  |||||  ::::|
Db 68576 ACGCATGAGTATATCAAGATTTATCAAGATCTGCTGCTGCTGCTGCTTAAAGATATTG 68635
OY 133 LeuGlyLysGlyIleAsnGlnAlaSerValGluIleSerPheGlyGluGluArgPro 152
    ::::|  |||||  ::::|  ::::|  ::::|  ::::|
Db 68636 GTATCAAAAGGTGCTGCTGCTGAACGTTTGAAACATCAAGGTGCAAGTTTGATTAACA 68695
OY 153 IleAlaPheGlyThrAsnGluGluAlaTyrSerGlnAsnArgArgAlaGlu 169
    ::::|  |||||  |||||  |||||  |||||  |||||
Db 68696 GTTGCATCAAAATGCTACCGAACAAGCTGCTTCCAAACCGCTGATTTAG 68746

```

RESULT 37

ABA91419 ID ABA91419 standard; DNA; 1035 BP.

AC ABA91419;

DT 18-APR-2002 (first entry)

DE Haemophilus paragallinarum strain 2403 haemagglutinin *hagA* gene.

KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;

KW *hagA*; gene; ds.

OS Haemophilus paragallinarum.

XX key location/Qualifiers

FT CDS 1..1035

FT sig_peptide 1..63

FT mat_peptide 64..1032

FT /tag= a

FT /product= "haemagglutinin"

FT /tag= b

FT /tag= c

XX MO200204485-A1.

PN 17-JAN-2002.

XX 06-JUL-2001; 2001WO-AU00822.

XX 07-JUL-2000; 2000AU-0008652.

XX (UYOU) UNIV QUEENSLAND.

XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;

XX WPI: 2002-154917/20.

XX P-PSDB; AAM50718.

PT New haemagglutinin polypeptide isolated from *Haemophilus paragallinarum*

PT useful as a vaccine for immunising chickens against coryza caused by

PT the species

PS Claim 11; Fig 5; 67pp; English.

XX The present sequence is that of the coding region of the *hagA* gene

CC of *Haemophilus paragallinarum* strain 2403 (serovar A). The gene

CC was isolated from chromosomal DNA by PCR amplification. It

CC encodes haemagglutinin (see AAM50718). The invention provides

CC recombinant haemagglutinin polypeptides (see AAM50716-27) and

CC A, B and C) of *H. paragallinarum*, the causative agent of infectious

CC coryza of chickens. The polypeptides are useful in vaccines for

CC immunisation against infectious bacteria, especially *Salmonella* or

CC when expressed in attenuated bacteria, especially *Salmonella* or

CC Mycoplasma (claimed). The recombinant polypeptide is preferably

CC the mature protein, or a biologically active fragment, variant or

CC derivative, that is capable of eliciting an immune response,

CC providing protection against one or more strains of *H.*

CC paragallinarum in chickens. Also claimed are methods of using the

CC haemagglutinin polypeptides and nucleic acids for detection and

CC diagnosis of infectious coryza in chickens.

XX Sequence 1035 BP; 319 A; 186 C; 234 G; 296 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5.71e-10	1035	51	25	68	21	5
Score:		161.50		46.06%		18.76%
Best Local Similarity:						
Query Match:						

US-09-674-779B-2 (1-172) x ABA91419 (1-1035)

```

OY 22 GlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyr 41
    |||  |||  |||||  |||||  |||||  |||||
Db 541 GGCAGTGTGGAAGAAATGATGAGCGGTGATTAATACCAAGATCGGTCTGTAACT 600
OY 42 ThrGlyValIleTyrThr-----GlyValAlaProLeuValAlaAspAspGluThrVal 59
    |||||  |||  |||||  |||||  |||||  |||||
Db 601 GCTGGTTTATCTTACCGCTTTGTGCAAGTGCACCAAGTTGACCTTAAGTTGTGCA 660
OY 60 LysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIle 79
    |||  |||  |||  |||  |||  |||  |||
Db 661 AAAACATTTGCA-----TTAAATTCAGATGTTACTTTCGATTTGGTAAACCAATTTA 714
OY 80 LysProGlnAlaAlaIleLeuAsp-----GluGlnAlaGlnPheLeuThr 95
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
Db 715 CGTCCGAAGCAACAAATGATTAAGACGATTTATGATGATGACACACATTA----- 768
OY 96 ThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlu 115
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
Db 769 -----AAATCACTACCAAGATGATGTTGCTGTTATACGACCGATTTGGTACGAAGCA 822
OY 116 TyrAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLys 135

```

```

Db      823 GCCAACTTGAATATATCAACACGCTGCTGATCTAGTCTAGTCTAGTCTAA 882
Qy      136 GYLILEASNGINALASERVALGILIELLESERPHGELYLUAGPROILEALPHE 155
Db      883 GGTGTGCTCAAGAGAGATTTCTTCAACAGGTATGTCAGAACCCAGTACTGCT 942
Qy      156 GLYTHRASNGLU-----GLUALATPSESGINASNARG 166
Db      943 GCGAATGTGATACGCTTAAAGTCGCAACGATTATCCCTGTTTAGCAGACGATCGT 1002
Qy      167 ARGALAGLUENUSER 171
Db      1003 CGGTAGAAATCTCA 1017

RESULT 38
ABA91422
ID      ABA91422 standard; DNA; 1035 BP.
AC      ABA91422;
XX      18-APR-2002 (first entry)
DE      Haemophilus paragallinarum strain 0222 haemagglutinin haga gene.
KW      Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;
XX      haga; gene; ds.
OS      Haemophilus paragallinarum.
FH      Key
FT      CDS
FT      Location/Qualifiers
FT      1..1035
FT      /*tag= a
FT      /product= "haemagglutinin"
FT      sig_peptide
FT      1..63
FT      /*tag= b
FT      mat_peptide
FT      64..1032
FT      /*tag= c

XX      WO200204485-A1.
XX      17-JAN-2002.
XX      PD
XX      PF      06-JUL-2001: 2001MO-AU00822.
XX      PR      07-JUL-2000: 2000AU-0008652.
XX      PA      (UYQU ) UNIT QUEENSLAND.
XX      PI      Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX      DR      MPI, 2002-154917/20.
XX      DR      P-PSDB: AAM50721.
XX      PS      New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX      PT      useful as a vaccine for immunising chickens against coryza caused by
XX      PT      the species
XX      PS      Claim 11: Fig 5; 67pp; English.

CC      The present sequence is that of the coding region of the haga gene
CC      of Haemophilus paragallinarum strain 0222 (serovar B). The gene
CC      was isolated from chromosomal DNA by PCR amplification. It
CC      encodes haemagglutinin (see AAM50721). The invention provides
CC      recombinant haemagglutinin polypeptides (see AAM50716-27) and
CC      encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
CC      A, B and C) of H. paragallinarum, the causative agent of infectious
CC      coryza of chickens. The polypeptides are useful in vaccines for
CC      immunisation against infectious coryza, as are the nucleic acids
CC      when expressed in attenuated bacteria, especially Salmonella or
CC      Mycoplasma (claimed). The recombinant polypeptide is preferably
CC      the mature protein, or a biologically active fragment, variant or
CC      derivative, that is capable of eliciting an immune response.

```

```

CC      providing protection against one or more strains of H.
CC      paragallinarum in chickens. Also claimed are methods of using the
CC      haemagglutinin polypeptides and nucleic acids for detection and
CC      diagnosis of infectious coryza in chickens.
XX      SO      Sequence 1035 BP; 318 A; 166 C; 234 G; 297 T; 0 other:

Alignment Scores:
Pred. No.:      5.71e-10      Length:      1035
Score:          161.50      Matches:      51
Percent Similarity: 46.06%      Conservative: 25
Best Local Similarity: 30.91%      Mismatches: 68
Query Match:    18.76%      Indels:      21
DB:             24      Gaps:        5

US-09-674-779b-2 (1-172) x ABA91422 (1-1035)
Qy      22 GLYCYSALASNLVSERTHRSEGLNVALLMETVALALAPROASNALAPROTHGLYTYR 41
Db      541 GGACGTGTGGAAGAAAGATGTRGCGGTGATGATTATACACGACATCGGTTCTGTAAT 600
Qy      42 ThrGlyValIleTyrThr-----GlyValAlaProLeuValAspAsnAspGluThrVal 59
Db      601 GCTGCTTATCTTACCGCTTTTGTCAAGTGCACACAGTTGTGAACCTAAGTTGTGCA 660
Qy      60 LysAlaLeuAlaSerLysLeuProSerLeuValIlyrPheAspPheAspSerAspGluIle 79
Db      661 AAACATTGCA-----TTAATTCAGATGCTTACTTCCGATTTGGTAAAGCAAAATTTA 714
Qy      80 LysProGlnAlaAlaAlaIleLeuAsp-----GluGlnAlaGlnPheLeuThr 95
Db      715 CGTCCAGAGACACAAATGATTATGACGCGTATTATGTTGAATCGCACGTTA----- 768
Qy      96 ThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlu 115
Db      769 -----AATACAGTACCAAGTAGATGTTGCTGTTATCTACTGACCCCTATTGGTAGGAGCA 822
Qy      116 TyrAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsnIlyrLeuGlyLys 135
Db      823 GCCAACTTGAATATATCAACACGCTGCTGATCTAGTCTAGTCTAGTCTAA 882
Qy      136 GYLILEASNGINALASERVALGILIELLESERPHGELYLUAGPROILEALPHE 155
Db      883 GGTGTGCTCAAGAGAGATTTCTTCAACAGGTATGTCAGAACCCAGTACTGCT 942
Qy      156 GLYTHRASNGLU-----GLUALATPSESGINASNARG 166
Db      943 GCGAATGTGATACGCTTAAAGTCGCAACGATTATCCCTGTTTAGCAGACGATCGT 1002
Qy      167 ARGALAGLUENUSER 171
Db      1003 CGGTAGAAATCTCA 1017

RESULT 39
ABA91423
ID      ABA91423 standard; DNA; 1035 BP.
XX      ABA91423;
XX      AC      ABA91423;
XX      DT      18-APR-2002 (first entry)
XX      DE      Haemophilus paragallinarum strain 2671 haemagglutinin haga gene.
XX      KW      Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;
XX      KW      haga; gene; ds.
XX      OS      Haemophilus paragallinarum.
XX      FH      Key
XX      FT      CDS
XX      FT      Location/Qualifiers
XX      FT      1..1035
XX      FT      /*tag= a
XX      FT      /product= "haemagglutinin"
XX      FT      sig_peptide
XX      FT      1..63

```




GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:16:20 ; Search time 1402 Seconds

(without alignments)
1986.893 Million cell updates/sec

Title: US-09-674-779b-2

Perfect score: 861

Sequence: 1 NMHIIQIAAAALSVLFEM.....IAFGTNEAMQSNRRALSY 172

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlj
-Q/cgcn2.1/USPTO.spool/US09674779/runat_06072003_121614_9813/app.query.fasta_1.327
-DB=EST -OFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -IOPTCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=100
-DOCALIGN=200 -THR.SCORE=dot -THR.MAX=100 -THR.MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09674779.cgcn2.1.1525 -@runat_06072003_121614_9813 -NCPU=6 -ICPU=3
-NO_MAP -LARGEOUTER -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEL_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_estlum:*
3: em_estlu:*
4: em_estlu:*
5: em_estlu:*
6: em_estlu:*
7: em_estlu:*
8: em_estlu:*
9: em_estlu:*
10: em_estlu:*
11: em_estlu:*
12: em_estlu:*
13: em_estlu:*
14: em_estlu:*
15: em_estlu:*
16: em_estlu:*
17: em_estlu:*
18: em_estlu:*
19: em_estlu:*
20: em_estlu:*
21: em_estlu:*
22: em_estlu:*
23: em_estlu:*
24: em_estlu:*
25: em_estlu:*
26: em_estlu:*
27: em_estlu:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	206	23.9	747	10	BE530264
2	182.5	21.2	279	17	BH392144
3	163	18.9	236	12	BE635668
4	163	18.9	246	12	BE651732
5	163	18.9	305	12	BE651173
6	161.5	18.8	567	17	BE459093
7	150	17.4	567	10	BE248360
8	130	15.1	801	17	BH404849
9	126	14.6	603	17	BH403497
10	124.5	14.5	814	17	BH386478
11	124	14.4	512	17	BH374389
12	121	14.1	658	17	BH387230
13	120	13.9	822	17	BH393516
14	117.5	13.6	743	17	BH389388
15	117	13.6	448	17	BH377553
16	113.5	13.2	800	17	BH376004
17	105.5	12.3	704	17	AQ012123
18	99	11.5	649	17	BH393822
19	93.5	10.9	296	9	A1904113
20	86.5	10.0	1152	14	BM802521
21	82.5	9.6	686	13	B1290925
22	82	9.5	638	17	BH388977
23	81.5	9.5	594	13	BM499884
24	81.5	9.5	930	14	BQ678932
25	81	9.4	565	10	BM448527
26	81	9.4	620	13	BE372739
27	80.5	9.3	1591	12	BE729857
28	80	9.3	458	12	BE758001
29	80	9.3	908	14	BQ218323
30	80	9.3	1010	13	BM475133
31	80	9.3	1131	13	BM1670158
32	79.5	9.2	385	17	AQ847628
33	79.5	9.2	480	10	BE135571
34	79.5	9.2	919	17	CNS07510
35	79	9.2	460	14	W20805
36	79	9.2	1083	12	BG261745
37	78.5	9.1	503	10	BE261626
38	78	9.1	360	14	D34784
39	78	9.1	676	13	AQ954996
40	78	9.1	676	13	BM491955
41	77.5	9.0	412	10	BE640026
42	77.5	9.0	457	17	CNS07ERR
43	77.5	9.0	578	13	B1595953
44	77.5	9.0	584	13	BM501746
45	77.5	9.0	662	14	BU028514
46	77.5	9.0	791	10	BE514244
47	77	8.9	413	14	D58855
48	77	8.9	551	12	BG559984
49	77	8.9	576	10	AV590145
50	77	8.9	595	13	B1066313
51	77	8.9	649	14	BQ991288
52	77	8.9	737	13	B1329425
53	77	8.9	877	13	B1957118
54	77	8.9	917	10	AW106318
55	77	8.9	958	13	BM475111
56	77	8.9	1747	17	BH770900
57	76.5	8.9	482	13	B1454957
58	76.5	8.9	559	13	BM499883
59	76.5	8.9	718	13	B1330404
60	76.5	8.9	773	13	B1333907
61	76.5	8.9	810	12	BF162117
62	76.5	8.9	840	13	B1554962
63	76.5	8.9	902	13	B1261918
64	76.5	8.9	905	17	CNS06001

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA

/clone_lib="Photobhabdus luminescens strain W14 M13 library"
 /dev_stage="primary phase variant"
 /note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janss."
 BASE COUNT 164 a 118 c 137 g 145 t 3 others
 ORIGIN

Alignment Scores:

Pred. No.: 2,51e-10 Length: 567
 Score: 161.50 Matches: 48
 Percent Similarity: 50.00% Conservative: 22
 Best Local Similarity: 34.29% Mismatches: 50
 Query Match: 18.76% Indels: 20
 DB: 17 Gaps: 4

US-09-674-779b-2 (1-172) x AQ990953 (1-567)

OY 1 MetwettLeuHISlegInIleAlaAlaAlaAlaAlaSerValLeuThrPheMet 20
 Db 198 CTGATGTTAGCTTACCAAGATTATGCGCTACAGCGTACTTCT----- 242
 OY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
 Db 243 -----AACAGAAAT-----GCTGATTAACGATCAAGCTGCT 272
 OY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnGluThrValLys 60
 Db 273 GTCGGCACTGTTAAAGAACTAATGCGGGT-----CTGTCAGACAGGAATTGCTGCT 326
 OY 61 AlaLeuAlaSerLysLeuPro-----SerLeuValTyrPheAspPheAspSerAspGlu 78
 Db 327 CAGCAATATGCAAGAGCTCAAAACAAACATCGTATATTGTTTGGACAATACGAC 386
 OY 79 IleLysProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAspGln 98
 Db 387 GTTGGCTCAGATTTTACTACAGCTGTATACGCTCAGCTCCTCCGCGACCAACCA 446
 OY 99 ThrAlaArgValLeuValAlaGlyHisThrAspGluArgLysSerArgLysTyrAspMet 118
 Db 447 TCTTTAAAGTTGTTGTGAAGAAGTCATGCTGACGACAGCTGTTCTCTGACTACACATC 506
 OY 119 SerLeuGlyGluArg-ArgAlaValAlaAlaArgAsnTyrLeuGlyLysGlyIle 137
 Db 507 GCATTANGTGAACGCCGCTGCTAATGACGTAATAATGACTGCAAGCAAAAGNGTT 564

RESULT 7

BE248360

LOCUS BE248360 237 bp mRNA linear EST 21-DEC-2000
 DEFINITION NF003G05DT1F1036 Drought Medicago truncatula cDNA clone NF003G05DT

ACCESSION BE248360

VERSION BE248360.2 GI:11964908

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 237)
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosida I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 Contact: May GD
 On Jul 13, 2000 this sequence version replaced gi:11964908.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7391
 Fax: 580 221 7380

Email: gdmay@noble.org
 Medicago Genome Initiative accession: MGI:S:22852
 Insert Length: 866 Std Error: 0.00
 Plate: 003 row: G column: 05
 Seq primer: TCACACAGCAAGACAGCTATGAC.
 Location/Qualifiers
 1.237
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="NF003G05DT"
 /clone_lib="Drought"
 /tissue_type="Plantlets"
 /dev_stage="Pooled timepoints"
 /note="Vector: Lambda Zap; Contains a mixture of entire plantlets harvested in a series of days-post-watering timepoints."

BASE COUNT 66 a 53 c 58 g 60 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.78e-09 Length: 237
 Score: 150.00 Matches: 31
 Percent Similarity: 74.14% Conservative: 12
 Best Local Similarity: 53.45% Mismatches: 14
 Query Match: 17.42% Indels: 1
 DB: 10 Gaps: 0

US-09-674-779b-2 (1-172) x BE248360 (1-237)

OY 116 TyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyrLeuGlyLys 135
 Db 3 TACACATCTCCCTGGGTGAGCTGCGAAGCCGCTTAAGATGACTGCAAGGTAA 62
 OY 136 GlyTleAsnGlnAlaSerValGluIleIleSerPhe-GlyGluGluArgProIleAla 155
 Db 63 GGGGTTTCGACAGACCGATCTCCATGCTTTCTTACAGGTAAAGAAACCTGCACT 122
 OY 155 eGlyThrAsnGluGluAlaTyrSerGlnAsnArgAlaGluLeuSerTyr 172
 Db 123 GGGTCATGACGAAGCGGCACTATCCAAACCGCTCGGCTAGCTGTTTAC 174

RESULT 8

BH404849/c

LOCUS BH404849 801 bp DNA linear GSS 11-DEC-2001
 DEFINITION AG-ND-126C17.TF ND-TM Anopheles gambiae genomic clone AG-ND-126C17

ACCESSION BH404849

VERSION BH404849.1 GI:17351065

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 801)
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 Contact: Brenden J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjl@utsl.tigr.org
 This clone is from an A. gambiae BAC library (ND-TM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center

University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1. 801

/organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone_id="ND-126C17"
 /note="Vector: pECBAC1; site_1: HindIII"

BASE COUNT 237 a 157 c 124 g 283 t

ORIGIN

Alignment Scores:

Pred. No.: 5.1e-06 Length: 801
 Score: 130.00 Matches: 27
 Percent Similarity: 55.21% Conservative: 26
 Best Local Similarity: 28.12% Mismatches: 41
 Query Match: 15.10% Indels: 2
 DB: 17 Gaps: 2

US-09-674-779b-2 (1-172) x BH404849 (1-801)

QY 55 AsnAspGluThrValIysAlaLeuAlaSerIysLeuProSerLeuValTyrPheAspPhe 74

DB 447 AACATCAGAGTTGTTACTGAGAGTTTCAGAGAGCTTGAAAGA---ATCAATTTGATTAA 391

QY 75 AspSerAspGluIleIysProGlnAlaAlaIleLeuAspGlnAlaGlnPheLeu 94

DB 390 AACTCAGACAGAAATCTTCTCTCTAATACGCTAATCAAGAGCTGCTGTAATATATAC 331

QY 95 ThrThrasnGlnThr---AlaArgValIleuValAlaGlyHisThrAspGluArgGlySer 113

DB 330 AACTCATCTGAATATGGCTTCTAAATATACAGTTGTAGAGAGCTGATACAGAGGTACA 271

QY 114 ArgGluTyrAsnMetSerLeuGlyIuArgArgAlaValAlaValArgAsnTyrLeuLeu 133

DB 270 GTTCAGATATACAGAAATATCTCAGAAAGAGCTAATATATAGTAAGAACTATTATAC 211

QY 134 GlyIysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyIleu 149

DB 210 CAGAGTGGAGTTCAATCTTCAGATTAGAGCTGTAAAGATGGTAAG 163

RESULT 9 BH403497 603 bp DNA linear GSS 11-DEC-2001

LOCUS AG-ND-160F6.TF ND-TAM Anopheles gambiae genomic clone AG-ND-160F6, DNA sequence.

ACCESSION BH403497

VERSION BH403497.1 GI:17349713

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 603)

AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL Unpublished (2001)

COMMENT Other GSSs: AG-ND-160F6.TF

Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b1loftus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain

DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1. 603

/organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone_id="AG-ND-160F6"
 /note="Vector: pECBAC1; site_1: HindIII"

BASE COUNT 208 a 81 c 151 g 163 t

ORIGIN

Alignment Scores:

Pred. No.: 1.04e-05 Length: 603
 Score: 126.00 Matches: 29
 Percent Similarity: 55.56% Conservative: 21
 Best Local Similarity: 32.22% Mismatches: 38
 Query Match: 14.63% Indels: 2
 DB: 17 Gaps: 2

US-09-674-779b-2 (1-172) x BH403497 (1-603)

QY 54 AsnAspGluThrValIysAlaLeuAlaSerIysLeuProSerLeuValTyrPheAsp 73

DB 228 GACATATACAGAGTTGCAATGAGATGAGAAACAGAAATTAATAAAT---GGTATTTTCAT 284

QY 74 PheAspSerAspGluIleIysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPhe 93

DB 285 TTTATATAAAGCAACATTAACAAATGACAGCTCAAGCTGATGAGTGCACACATT 344

QY 94 LeuThrTrAsnGlnThrAlaArgValIleuValAlaGlyHisThrAspGluArgGlySer 113

DB 345 ATTAAGACCAAT---GGTGAATATTAATTAACCGACATACAGATGCTAAAGAGT 401

QY 114 ArgGluTyrAsnMetSerLeuGlyIuArgArgAlaValAlaValArgAsnTyrLeuLeu 133

DB 402 GCAGCATATCAACCTTAAGCTTCCAGAGAAAGAGCGCTGCTGTGGAGAGCTGGAA 461

QY 134 GlyIysGlyIleAsnGlnAlaSerValGlu 143

DB 462 GCTAGAGAGAGTTCTCCAGATCTTAA 491

RESULT 10 BH386478 814 bp DNA linear GSS 10-DEC-2001

LOCUS AG-ND-170A18.TF ND-TAM Anopheles gambiae genomic clone AG-ND-170A18, DNA sequence.

ACCESSION BH386478

VERSION BH386478.1 GI:17332620

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 814)

AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL Unpublished (2001)

COMMENT Other GSSs: AG-ND-170A18.TF

Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 658)	Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J	Direct Submission of BAC-end sequences from <i>Anopheles gambiae</i>	Unpublished (2001)	Other_GSSS: AG-ND-169L12.TR

FEATURES

Location/Qualifiers

BASE COUNT	260 a	86 c	132 g	180 t
ORIGIN				

Alignment Scores:	
Pred. No.:	5,31e-05
Score:	121.00
Percent Similarity:	44.78%
Best Local Similarity:	25.37%
Query Match:	14.05%
DB:	17
Length:	656
Matches:	34
Conservative:	26
Mismatches:	67
Indels:	2
Gaps:	2

US-09-674-779B-2 (1-172) x BH387230 (1-658)

QY	22	CysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThr	42
Db	201	TGCTGGACAGCAAGAAAGTGAAAAATTA	261
QY	43	GlyVal---TLeTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAla	61
Db	261	AAGGTGAAAAAATTA	320
QY	62	LeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLysPro	81
Db	321	GATTGGAAAT-----ATTAATTTTGATAATATAAAGCTACTCTTACAGCA	365
QY	82	GlnAlaAlaAlaLLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArg	101
Db	366	GATGGCAAAATTTGTGGCAGATGAGATCCTAAAGCTGTGGATGAAAGAAACCTCGAATTTAAA	423
QY	102	ValLeuValAlaGlyHisThrAspGluArgLysSerArgGluTyrAsn-MetSerLeuGlu	121
Db	426	CTTTATATGAAAGGGCATCTGACAAACAATGTCATAAAGAACAACATATGGAATATATC	485
QY	121	YgluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyLysGlnAlaLase	141
Db	486	ATTGACAGAGCTACACAACTATATCTTATTTAAACATATAAGGATTAATCTGATAG	545
QY	141	rValGluLleIleSerPheGlyGluGluArgProLleAla	154
Db	546	ACTACAAACAAAGGTTTGGACAGGATTAATCTTTGGCT	585

RESULT	13
LOCUS	BH393516
DEFINITION	BH393516 822 bp DNA linear GSS 11-DEC-2001 AG-ND-16ON3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-16ON3, DNA sequence.
ACCESSION	BH393516
VERSION	BH393516.1 GI:17339657
KEYWORDS	GSS. African malaria mosquito.
SOURCE	Anopheles gambiae

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 822)	Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J	Direct Submission of BAC-end sequences from Anopheles gambiae	Unpublished (2001)	
	Other_GSSs: AG-ND-160N3.TF			

FEATURES	Location/Qualifiers
source	1. .822

BASE COUNT	297 a	104 c	178 g	243 t
------------	-------	-------	-------	-------

Alignment Scores:	
Pred. No.:	0.000103
Length:	822
Score:	120.00
Percent Similarity:	53.33%
Best Local Similarity:	33.33%
Query Match:	13.94%
DB:	17
Length:	822
Matches:	30
Conservative:	18
Mismatches:	38
Indels:	3
Gaps:	2

US-09-674-779B-2 (1-172) x BH393516 (1-822)

QY	5	GLuTHrVallyAlAlenulaserlysluPser-----LeuVallyPhasp	73
Db	227	AMACATATACCAGGTTGCACACTGAGATTAGAACACAGATTAAAAATGTATTTTCAT	268
QY	74	phaspSerAspGluuileysProGlnAlaAlaileuaspgluGlnAlaGlnPhe	93
Db	287	TTTAATTAAGCAACATATAACAATATAGTCAGGTGGAAGCTAGATGTATGCCACAT	346
QY	94	LeuTHrThAsnGlnThrAlaArgValLeuValAlaGlyHsrthaspGluuArglySer	113
Db	347	ATTAAAGACCAAT---GCTGGAATATATCTATTAAACCCGACATACAGATGCTAAAGCTAGT	403
QY	114	ATrgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyrLeuLeu	133
Db	404	GCAGCATACAACTTAACTTCCAGAGAAGAAGCGCGCTGCTGTTGGAGCTCTGGAA	463

OY 134 GlytysgilyleasnglnalaserValglu 143
 DB 464 GCTAGAGGAGTTTCTTCAGTACTTAA 493

RESULT 14
 BH389388 743 bp DNA linear GSS 11-DEC-2001
 LOCUS AG-ND-132E3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-132E3,
 DEFINITION DNA sequence.
 ACCESSION BH389388
 VERSION BH389388.1 GI:17335529
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.

REFERENCE 1 (bases 1 to 743)
 AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other_GSSs: AG-ND-132E3.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjlloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 source location/Qualifiers
 1..743
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-132E3"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 269 a 100 c 169 g 205 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.000183 Length: 743
 Score: 117.50 Matches: 26
 Percent Similarity: 55.41% Conservative: 15
 Best Local Similarity: 35.14% Mismatches: 32
 Query Match: 13.65% Indels: 1
 DB: 17 Gaps: 1

US-09-674-779b-2 (1-172) x BH389388 (1-743)

OY 70 ValtyrPheaspPheaspSeraspSclulleySPROGlnAlaAlaIleLeuaspGlu 89
 DB 275 GTGATATTCATTATTAATTAAGCAACAATACAAATAGTCGCAAGCTGATGTA 334

OY 90 GlnAlaGlnPheLeuThrThraSnglnThraAlaArgValLeuValAlaGlyHisThraSp 109
 DB 335 GCTGCCAACATTATTAAAGCAAT--GGTGAATATTATTATTAACCGGACATACAGAT 391

OY 110 GluArgGlySerArgGluTyraSmetSerLeuGlyGluArgAlaValAlaValArg 129
 DB 392 GCTAAAGTATGTCAGCATACACCTTAAGCTTTCCAGACAAAGAGCGGCTGCTGTGTT 451

OY 130 AsnTyrlLeuGlyLeuSnglnAlaSerValglu 143
 DB 452 GCAGCTCTGAGACTAGAGAGTTTCTTCAGTACTTAA 493

RESULT 15
 BH377553 448 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-155H3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-155H3,
 DEFINITION DNA sequence.
 ACCESSION BH377553
 VERSION BH377553.1 GI:17323695
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.

REFERENCE 1 (bases 1 to 448)
 AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other_GSSs: AG-ND-155H3.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjlloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 For
 Class: BAC ends.

FEATURES
 source location/Qualifiers
 1..448
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-155H3"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 167 a 71 c 96 g 114 t
 ORIGIN

Alignment Scores:
 Pred. No.: 9.17e-05 Length: 448
 Score: 117.00 Matches: 33
 Percent Similarity: 52.94% Conservative: 12
 Best Local Similarity: 38.82% Mismatches: 36
 Query Match: 13.59% Indels: 4
 DB: 17 Gaps: 3

US-09-674-779b-2 (1-172) x BH377553 (1-448)

OY 87 LeuaspGlnGlnAlaGlnPheLeuThrThraSnglnThraAlaArgValLeuValAlaGly 106
 DB 39 CTGACAGAGTTGGCGCA---CTTAGATATATATGCAATTCAGATTTT-----GGA 89

OY 107 HisThraSpGlnArgGlySerArgGluTyraSmetSerLeuGlyGluArgAlaVal 126
 DB 90 AATTGGCAATTTCCGCAACTTTGCAATTAATAAAAGCTGTCGAAATCGCCCAAT 149

OY 127 AlaValArgAsnTyrlLeuGlyLeuSnglnAlaSerValglu---IleIle 145
 DB 150 GCGGTTAACGATATCTCAAGTCACAAATAATGAAACAAATATCAAACTGGAAGTGTGTA 209


```

Db      269  CGTACCTGACCGCCCTGCGCATGCGCCGAGACGCGTCCGCGCGTGGGATGGC 328
      |||||  |||  |||||  |||  |||
Qy      149  GluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgArgAla 168
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      329  GAACGG-----TTCTGTGGCCGACAGAAAGACCCCGCAGCGCGAAGAACCGGGGCTC 382
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      169  Glu 169
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      383  GAG 385

RESULT 18
BH393822
LOCUS   AG-ND-169D8.TF ND-TAM Anopheles gambiae genomic clone AG-ND-169D8,
DEFINITION
DNA sequence.
ACCESSION   BH393822
VERSION     BH393822.1  GI:17339963
KEYWORDS    GSS:
SOURCE      African malaria mosquito.
ORGANISM    Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
            Anophelinae.
REFERENCE   1 (bases 1 to 649)
AUTHORS    Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE      Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL     Unpublished (2001)
COMMENT     Other GSSs: AG-ND-169D8.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: b.loftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M University BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.
FEATURES
    source          location/Qualifiers
    1..649
    /organism="Anopheles gambiae"
    /strain="PST"
    /db_xref="taxon:7165"
    /clone_lib="AG-ND-169D8"
    /clone_11b="ND-TAM"
    /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      225 a      89 c      151 g      184 t
ORIGIN
Alignment Scores:
    Pred. No.:      0.0354      Length:      649
    Score:          99.00      Matches:      24
    Percent Similarity: 54.93%      Conservative: 15
    Best Local Similarity: 33.80%      Mismatches: 30
    Query Match:    11.50%      Indels:      3
    DB:            17      Gaps:      2
US-09-674-779B-2 (1-172) x BH393822 (1-649)
Qy      54  AspaenaspGluThrValIleuAlaLeuAlaSerIleuProSerLeuValTyrPheasp 73
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      234  GACATTAACCAAGGTTCGCAACTGAAAGTACAAACAGATTAAAAAT--GTTATTTCAT 290
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      74  PheaspSeraspGluIleuIleuProGluAlaIleuAlaIleuAspGluGluAlaGlnPhe 93

```

```

Db      291  TTTAATAAAGCAACATACATAAATAATGATGAGGCAAGCTAGATGATAGCTCCAAACAT 350
      |||||  |||  |||||  |||  |||
Qy      94  LeuThrThrAsnGlnThrAlaArgValIleuValAlaGlyHisThrAspGluArgGlySer 113
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      351  ATTAAGAGCAAT--GGTGGAAATTAATCTTAACCGGACATACAGATCTTAAGGTAGT 407
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      114  ArgGluTyrAsnMetSerLeuGlyGluArgArg 124
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      408  GCAGCATACACACTTAAGCTTTTCACACAAAGA 440

RESULT 19
AI904113/c
LOCUS   AI904113
DEFINITION
IL-BP042-231198-008 BT042 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AI904113
VERSION     AI904113.1  GI:6494500
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 296)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
            20202663
            Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/seq/gethtml.pl?tl=IL-BT042-008.html
            673-231198&t4=1)
            Seq primer: puc 18 forward.
FEATURES
    source          location/Qualifiers
    1..296
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="BT042"
    /sex="female"
    /dev_stage="Adult"
    /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
            SmaI; A mini-library was made by cloning products derived
            from ORESTES PCR (U.S. Letters Patent application No. 196
            716 - Ludwig Institute for Cancer Research) profiles
            into the puc 18 vector. Reverse transcription of tissue
            mRNA and cDNA amplification were performed under low
            stringency conditions."
BASE COUNT      43 a      88 c      97 g      68 t
ORIGIN
Alignment Scores:
    Pred. No.:      0.0491      Length:      296
    Score:          93.50      Matches:      21
    Percent Similarity: 48.19%      Conservative: 19
    Best Local Similarity: 25.30%      Mismatches: 36
    Query Match:    10.86%      Indels:      8
    DB:            9      Gaps:      1
US-09-674-779B-2 (1-172) x AI904113 (1-296)

```

QY 94 LeuThrasnglnThraAlaArgValleuAlaGlyHsthraspgluArgGlySer 113
 ||| ||| : : : : : ||| ||| : : :
 Db 284 CTTACACACCCCAACGCGGCAAGATGATATACCCG-CATGCCACACAGATGGCGAG 226
 QY 114 ArgGluThrasnMetSerLeuGlyGluArgAlaValAlaValArgsnTrileuLeu 133
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 225 CCGCAGAGCAACCGGCTTTCGCGCAACGTGCGAGACCATCAAGACCTACTGCTG 166
 QY 134 GlyLysGlyLeuAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgProIle 153
 ||| ||| : : : : : ||| : : : : : ||| : : : : :
 Db 165 GGCAGAGGCGCTCGCGCGGCAACTGTGTGATCCGTGGCGAGAGTAGCAACAACCCCTG 106
 QY 154 AlaPheGlyThrasnglu-----GluAlaTrpSerGlnAsnArg 166
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 105 GTGGCTGCAACATGCAAGCGCGCGGACCGCACTGATTCAGTCCGTGCAACTACAGG 46
 QY 167 ArgAlaGlu 169
 ||| ||| : : : : : ||| : : : : : ||| : : : : :
 Db 45 CGCCTCGAG 37

RESULT 20
 BM802521/c 1152 bp mRNA linear EST 05-MAR-2002
 LOCUS BM802521
 DEFINITION AGENCOURT_6460224 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5560157
 5', mRNA sequence.
 ACCESSION BM802521
 VERSION BM802521.1 GI:19119344
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 JOURNAL NIH-MGC http://mgs.nci.nih.gov/
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs@email.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LINL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.lnl.gov
 Plate: LHAM1285 row: 0 column: 06
 High quality sequence start: 379
 High quality sequence stop: 585.
 Location/Qualifiers

FEATURES

source
 1. 1152
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5560157"
 /clone_id="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-SPORT6;
 site:1: NotI; site:2: SalI; Cloned unidirectionally;
 oligo-dT primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH-MGC Library."
 BASE COUNT 344 a 277 c 241 g 273 t 17 others
 ORIGIN

Alignment Scores:

Pred. No.: 3.75 Length: 1152
 Score: 86.50 Matches: 30
 Percent Similarity: 48.33% Conservative: 28
 Best Local Similarity: 25.00% Mismatches: 43
 Query Match: 10.05% Indels: 19
 DB: 14 Gaps: 5

US-09-674-779b-2 (1-172) x BM802521 (1-1152)

QY 52 LeuValAspAsnAspGluThrValIleAlaLeuAlaSerIleuProSerLeuValTyr 71
 ||| ||| : : : : : ||| ||| : : : : : ||| : : : : :
 Db 785 TTGGTCAATCAATCAACCATTTATTAAGTCCCTGGCGCAATATATCAAGATGCTTC 726
 QY 72 PheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGluGlnAla 91
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 725 ATTCATCAGAGAAAGAT-----CCATCAGATGCTGGCGCAAGTTCTTCA 678
 QY 92 GlnPheLeuThrThrasnglnThraAlaArgValleuAlaGlyHsthraspgluArg 111
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 677 ACAATGGCGACCTCTTCGGGAATCTCGCCT-----CCTGCTCTTCGCGAG 627
 QY 112 GlySerArgGluThrasnMetSerLeuGlyGluArgAlaValAlaValArgsnTrileu 131
 ||| ||| : : : : : ||| : : : : : ||| : : : : :
 Db 626 CCACCAATGAAATTCCTTGGCAGCTG-----GCCCTCCAAATAGACCTTAC 579
 QY 132 LeuLeuGlyLysGlyLeuAsnGlnAlaSerValGlu-----IleIleSerPheGlyGlu 150
 : : : : : ||| : : : : : ||| : : : : :
 Db 578 ATCACTGGTAAAGTCCCTTCAGCATATCTCAATTCATGTTAGCTTGGCAATGTT 519
 QY 151 ArgProIleAlaPheGlyThrasnglu-----GluAlaTrpSer 163
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 518 ATTCCTGTCCTTGTGATTCATCAACCTCTGCAACAATCCTTGGCAGCGCTGAGT 459

RESULT 21
 BJ290925/c 686 bp mRNA linear EST 09-APR-2002
 LOCUS BJ290925
 DEFINITION BJ290925 Y. Ogihara unpublished cDNA library, Wh_SL Trilicium
 aestivum cDNA clone whs120n03 5', mRNA sequence.
 ACCESSION BJ290925
 VERSION BJ290925.1 GI:20108010
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Trilicium aestivum

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 JOURNAL 1 (bases 1 to 686)
 COMMENT 1 (bases 1 to 686)
 Ogihara, Y. and Mural, K.
 Expressed genes in Trilicium aestivum
 Unpublished (2002)
 Contact: Tadasu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers

FEATURES

source
 1. 686
 /organism="Trilicium aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="whs120n03"
 /clone_id="Y. Ogihara unpublished cDNA library, Wh_SL"
 /tissue_type="seed DPA30"
 /dev_stage="Feekes' scale 11.3"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site:1: EcoRI; Site:2: XhoI; Plants were grown under
 hydroponic conditions at UC Davis, salt stressed for 12
 hours, and for 7 days, then dissected and frozen (Akhunov
 in J. Dvorak Lab). Total RNA was prepared from sheath
 tissue, equal quantities of RNA were pooled from the two
 samples, polyA was purified from the pooled RNA, a cDNA
 library was made, and the cDNA clones were in vivo
 excised to give plasmid phagemids in the T7 Close lab
 at the University of California, Riverside (Akhunov, Chin
 , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang).
 Plasmid DNA preparations and DNA sequencing were
 performed in the OD Anderson lab (all other authors)."
 BASE COUNT 138 a 233 c 181 g 132 t 2 others
 ORIGIN

COMMENT

Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 952035 row: D column: 03.

FEATURES

SOURCE

1. 594
/organism="Zea mays"
/cultivar="BMS (Black Mexican Sweet)"
/db_xref="taxon:4577"
/clone_lib="952 - BMS tissue from Walbot Lab (reduced RNA)"
/tissue_type="suspension culture"
/dev_stage="mixed logarithmic and stationary growth phases"
/lab_host="DH10B"
/note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive clones."

BASE COUNT 177 a 81 c 202 g 134 t
ORIGIN

Alignment Scores:

Pred. No.: 5.49 Length: 594
Score: 81.50 Matches: 35
Percent Similarity: 40.94% Conservative: 17
Best Local Similarity: 27.56% Mismatches: 33
Query Match: 9.47% Indels: 42
DB: 13 Gaps: 7

US-09-674-779b-2 (1-172) x BM499884 (1-594)

QY 46 TyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLys 65
DB 288 TGAATGATGCTCTCACTGATGTTGAAGAGAGAGAG----- 323
QY 66 LeuProSerLeuValTyrPheAspPheAspSerAspGluLeuLysProGlnAlaAlaAla 85
DB 324 -----TTGGATTGTTGATGAGATGAGCCAAAGCCAAAGAGAGAGCA 365
QY 86 IleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAla 105
DB 366 TTGCTGATGACAACTTGTGCTTCTTGGAGAA-----TAT 407
QY 106 GlyHisThrAspGlu-----ArgGlySerArgGluTyr-----Asn 117
DB 408 GCACACAGTATGATGATGACATGCTGTTAAAGCTGG--GAATGCAACTGCCAAAT 464
QY 118 MetSerLeuGlyGluArgAlaVal-AlaValArgAsnTyrLeuLeuGlyLysGly 137
DB 465 GAGCTATGATGACAAATGCTCCAAATCAGATGTATGTGTGATGAGAGATAT 524
QY 137 eAsnGlnAlaSerValGluLeuLeuSerPheGlyGluGluArgProIleAlaPheGlyThr 157
DB 525 A-----GAACACCAGCAGATTTTGTTCGCCAA 551
QY 157 Asn---GluGlnAlaTyr 162
DB 552 AACTGATTCACATGAG 570
RESULT 24
BQ678932
LOCUS BQ678932 930 bp mRNA linear EST 15-JUL-2002

DEFINITION

AGENCOURT 8192156 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6260167 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
human.
NIH-MGC http://mgs.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgs@bbs.fremail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at:
http://image.llnl.gov
Plate: LRCM2418 row: n column: 08
High quality sequence stop: 714.

FEATURES

SOURCE

1. 930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="6260167"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGCAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

BASE COUNT 266 a 283 c 186 g 193 t 2 others
ORIGIN

Alignment Scores:

Pred. No.: 11.6 Length: 930
Score: 81.50 Matches: 38
Percent Similarity: 42.03% Conservative: 20
Best Local Similarity: 27.54% Mismatches: 36
Query Match: 9.47% Indels: 44
DB: 14 Gaps: 8

US-09-674-779b-2 (1-172) x BQ678932 (1-930)

QY 27 SerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyr 46
DB 470 AGCAGACGACAGCAGCTGCTCCAGAGCCTGTGGT----- 511
QY 47 ThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeu 66
DB 512 ACAGATTTCAAACCTTA-----AATGATACCCCTGCACACAGACAGACAGACC 559
QY 67 ProSerLeuValTyrPheAspPhe-----Asp 75
DB 560 CCAAGGCCACATTCCTGCTTATACACAGCTACAGCTCAACAACTAGTACAAACAAT 619
QY 76 SerAspGluLeuLysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThr 95
DB 620 AGTACTGACAGCTTAACCA---GGGCTTCAATTAACAGTAAAGCTGCTCACTTACACACA 676
QY 96 ThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlu 115
DB 677 ACTAGTGCACCACTAAGTTGATC-----CATCCAGATGAG----- 712

OY 116 TyranMetSerIeuGlyuArgAlaValAlaValArgAsn-TyrIeuIeuGlyLy 135
 Db 713 ---GATATATCCCTGGAGAGAGAAGGCA-----CAGTTACCTA----- 749
 OY 135 sGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluArgPro 152
 Db 750 -----AGTATCAACGTAACTTCTCGCCGAGACAGAGCC 785

RESULT 25
 BM448527 565 bp mRNA linear EST 01-APR-2002
 LOCUS DS4025B03.59675 An expressed sequence tag database for the
 DEFINITION halotolerant green alga, Dunaliella salina Dunaliella salina cDNA
 clone DS4025B03 5, mRNA sequence.
 ACCESSION BM448527
 VERSION BM448527.1 GI:19854099
 KEYWORDS EST.
 SOURCE Dunaliella salina.
 ORGANISM Dunaliella salina.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Dunaliellaceae; Dunaliella.
 REFERENCE 1 (bases 1 to 565)
 AUTHORS Cushman,J.C.
 TITLE An expressed sequence tag database for the halotolerant green alga,
 Dunaliella salina
 JOURNAL Unpublished (2002)
 COMMENT Contact: Cushman JC
 Department of Biochemistry
 University of Nevada
 MS200, Reno, NV 89557-0014, USA
 Tel: 775-784-1918
 Fax: 775-784-1650
 Email: jcushman@unr.edu
 PCR Primers
 FORWARD: T3 20mer
 BACKWARD: T7 21mer
 Plate: 025 row: B column: 03
 Seq primer: T3 20mer
 High quality sequence stop: 565.
 Location/Qualifiers
 1..565
 /organism="Dunaliella salina"
 /db_xref="taxon:3046"
 /clone="DS4025B03"
 /clone_lib="An expressed sequence tag database for the
 halotolerant green alga, Dunaliella salina"
 /tissue_type="Cells, which was adapted in 2.5M NaCl via a
 incremental series from 1.7 to 2.0 to 2.25 to 2.5 M NaCl,
 were exposed to 3.4 M NaCl for 5 hours"
 /cell_type="Green"
 /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site.1:
 EcoRI; Site.2: XhoI. Library construction was performed
 according to Stratagene's recommended protocol for the
 Lambda UniZapXR vector and cDNA synthesis kit."
 BASE COUNT 131 a 163 c 179 g 92 t
 ORIGIN

Alignment Scores:
 Pred. No.: 5.86 Length: 565
 Score: 81.00 Matches: 36
 Percent Similarity: 41.33% Conservative: 26
 Best Local Similarity: 24.00% Mismatches: 54
 Query Match: 9.41% Indels: 34
 DB: 13 Gaps: 5

US-09-674-779b-2 (1-172) x BM448527 (1-565)

OY 9 AlaAlaAlaAlaAlaAlaSerValIeuThrPheMetThrGlyCysAlaAsnLysSerThr 28
 Db 11 GCAGAGCCCATGAAGCTTAACATCTTACCCGCCACCTGGCTCCCAAGAAAGACCTGAG 70
 OY 29 SerGlnValMetValAlaIleProAsnAlaProThrGlnIleTyrThrGlyValIleTyrThrGly 48

Db 70 ----- 70
 OY 49 ValAlaProIeuValAlaSpasnaSpGluThrValLysAlaLeuAlaSerLys---IeuPro 67
 Db 71 -----ATCATATATAGACGCCAAGCTGCGGCTTCTACGACAAAGCGCTGCC 118
 OY 68 SerIeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleIleu 87
 Db 119 GCAGAGCTGGAGGGCCAGGCCCTTGGCCACAGATTCAGAGGATTTTCAAGATCATG 178
 OY 88 -----AspGluGlnAla-----GlnPheIeuThrThrAsnGlnThrAla 100
 Db 179 GCAGACAGACAGACAGACAGAGGTTTCCGACATGAAGAGGAGGTGCTCACCACCAACGCTGTG 238
 OY 101 ArgValIeuValAlaGluIleThrAspGluArgGlySerArgGluTyrAsnMetSerIeu 120
 Db 239 CGCTGCTCATGTGCC-----CCCGAGACCAAGCTTCAGGGGTGATGCGCGCTGAT 292
 OY 121 GlyIuArgArgAlaValAlaValAlaValArgAsnTyrIeuGlyLysGlyIleAsnGlnAla 140
 Db 293 GCGAGCGAGCGAGCGAGAGAGTGTGCGCGCTGCTGATGCTGCTGCTGCTGCTGCTG 352
 OY 141 SerValGluIleIleSerPheGlyGluIu 150
 Db 353 AACCTGTGATGCTGTAAGAAGGCGAGCAG 382

RESULT 26
 BE372739 620 bp mRNA linear EST 21-JUL-2000
 LOCUS 601224250F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3582660 5',
 DEFINITION mRNA sequence.
 ACCESSION BE372739
 VERSION BE372739.1 GI:9318102
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 620)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.jnl.gov
 Plate: L1A8738 row: C column: 13
 High quality sequence stop: 599.
 Location/Qualifiers
 1..620
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3582660"
 /clone_lib="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site.1: SalI;
 Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 137 a 179 c 175 g 129 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6.84 Length: 620
 Score: 81.00 Matches: 41

QY 111 gGlySerArgGlu 115
 Db 385 AGCCAGAGAGAG 397

RESULT 30
 LOCUS BM475133/c 1010 bp mRNA linear EST 05-FEB-2002
 DEFINITION AGENCOURT_6477419 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5559273
 5', mRNA sequence.
 ACCESSION BM475133
 VERSION BM475133.1 GI:18524175
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 1010)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM12283 row: 3 column: 10
 High quality sequence stop: 517.

FEATURES
 Source
 1..1010
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5559273"
 /clone_lib="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-SPORT6;
 Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
 oligo-dt primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 316 a 224 c 214 g 256 t

ORIGIN

Alignment Scores:
 Pred. No.: 20.7 Length: 1010
 Score: 80.00 Matches: 34
 Percent Similarity: 47.33% Conservative: 28
 Best Local Similarity: 25.95% Mismatches: 55
 Query Match: 9.29% Indels: 14
 Db: 13 Gaps: 5

US-09-674-779b-2 (1-172) x BM475133 (1-1010)

QY 30 GluValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGlyVal 49
 Db 671 CAGGTGATGTTGTGAGAGAGATCTGAAATAGTGATATTTCTCTGCAGAGATT 612

QY 50 AlaProLeuValAspAsnAspGluThrValValAlaLeuAlaSerLysLeuProSerLeu 69
 Db 611 -----TTGGTCTGAATCAACCCAGTTTAAAGTCCGCGGCAAGATTATCAAGAATG 558

QY 70 ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGlu 89
 Db 557 GTCTCATCTTCATCAGAGAAAGAACATCAGCATCTGCGGCAAGTTCTTCAACAT 498

QY 90 GlnAlaGlnPheLeuThrAsnGlnThrAlaValIleValAlaGlyHisThrAsp 109
 Db 497 GCGACCTCTTCTGGGAATCTCGCTCTCTCG-----GCCTTTCAC----- 456

QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaArg 129
 Db 455 ---CAGCAAGCAATTAATTCCTGGACCTTG-----GCCTTCCAAATTAAGA 411

QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGlu---IleIleSerPheGly 148
 Db 410 ACTTACATCATCTGTAAGAGTCCCTCAGACATCTCTTAATTCATCTGACGTTTGCC 351

QY 149 GluGluArgProIleAlaPheGlyThrAsnGlu 159
 Db 350 AATGTTATTCCTGTTCTCTCTGTTATTCATCA 318

RESULT 31
 LOCUS BI670158/c 1131 bp mRNA linear EST 12-SEP-2001
 DEFINITION 603294002F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:531336 5',
 mRNA sequence.
 ACCESSION BI670158
 VERSION BI670158
 KEYWORDS BI670158.1 GI:15584391
 EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 1131)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM11794 row: 3 column: 01
 High quality sequence stop: 4.

FEATURES
 source
 1..1131
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:531336"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTCTTTTCTTTTCTTTV-3',
 size selected for average insert size 2.3 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 242 a 277 c 359 g 253 t

ORIGIN

Alignment Scores:
 Pred. No.: 25 Length: 1131
 Score: 80.00 Matches: 35
 Percent Similarity: 41.74% Conservative: 13
 Best Local Similarity: 30.43% Mismatches: 49
 Query Match: 9.29% Indels: 18
 Db: 13 Gaps: 4

US-09-674-779b-2 (1-172) x BI670158 (1-1131)

QY 29 SerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGly 48
 Db 455 -----TTGGTCTGAATCAACCCAGTTTAAAGTCCGCGGCAAGATTATCAAGAATG 558

Db 336 TCCCAATACATG-----TCCCGGGGGGTCCGAGGAGTCTCTTACACTCT 289

Qy 49 ValAlaProLeuValAspAsnAspGluThrVal-----LysAlaLeuAlaSerLysLeu 66

Db 288 GTGACATTATTCATTAAGACAGAAAGACGACCACTCTCAGACCACTCTCTG 229

Qy 67 ProSerLeuValTyrPheAspPheAspSerAspGluLeuLysProGlnAlaAlaAla 86

Db 228 CCCCTGCCCTAGACTCTCTTACACCGGCTCTACTGTAGACGACGACGACCTT 169

Qy 87 LeuAspGluGlnAlaGlnPheLeuThrThrasnGlnThrAlaArgValLeuAlaGly 106

Db 168 TCGATGAGTCTAGACCTCTCTCAGTCCGCGCAGCCGACAGCAGTC----- 121

Qy 107 HisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGluArgAlaVal 126

Db 120 -----CATGATCTCGAAGCTCTCTGCTTCA-----CGAGGTG 85

Qy 127 AlAlaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSer 141

Db 84 GCCCTCCCTCAGCTCCGCTCGAGCATGAGTGTGCTGCTCATCG 40

RESULT 32

LOCUS A0847628 385 bp DNA linear GSS 25-MAY-2001

DEFINITION LMAJFV1_lm35c11.y1 Leishmania major FV1 random genomic library

ACCESSION A0847628

VERSION A0847628.1 GI:6052276

KEYWORDS GSS.

SOURCE Leishmania major

ORGANISM Leishmania major

REFERENCE 1 (bases 1 to 385)

AUTHORS Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L., Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Blistein, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagaris, W., Williams, T., Jackson, Y., Bowers, Y., Swaller, T., Waterston, R., Wilson, R., and Beverley, S.M. A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing: a resource for DNA microarrays and expression profiling

TITLE Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)

JOURNAL 21192569

MEDLINE Contact: Akopyants, NS / Beverley, SM

COMMENT Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center

If using this information please cite:

N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing' and the Washington University Genome Sequencing Center for information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@borcim.wustl.edu)

Seq primer: -40RP from Gibco

Class: shotgun

High quality sequence stop: 365.

Location/Qualifiers

1..385

/organism="Leishmania major"

/strain="Friedlin strain VI"

/db_xref="taxon:5664"

/clone="LMAJFV1_lm35c11"

/clone_lib="Leishmania major FV1 random genomic library"

/lab_host="TOP10 (Invitrogen)"

/note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV; Genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNA polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 97 a 120 c 115 g 53 t

ORIGIN

Alignment Scores:

Pred. No.:	4.83	Length:	385
Score:	79.50	Matches:	34
Percent Similarity:	38.89%	Conservative:	15
Best Local Similarity:	26.98%	Mismatches:	44
Query Match:	9.23%	Indels:	33
DB:	17	Gaps:	5

US-09-674-779b-2 (1-172) x A0847628 (1-385)

Qy 34 AlAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGlyValAlaProLeuVal 53

Db 8 GCCCCAGGACGCCCGCGGAGGTGACCGGACGCTCATCACCCTCAGGAACTCGCT 67

Qy 54 AspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheasp 73

Db 68 -----GCTTGAAGCAGAGGCGCGTGGAGGAGGAGGAGC 100

Qy 74 PheAspSerAspGluIleLysProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPhe 93

Db 101 TTGTGATGCCGAGGAGGACTCAAGCG-----CAGATC 133

Qy 94 LeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgLysSer 113

Db 134 ATCACAATC-----GAGAACGCCGAG 154

Qy 114 ArgGluTyr---AsnMetSerLeuGlyGluArgAlaValAlaValAlaArgAsnTyrLeu 132

Db 155 CGCGAGGTTGCACAGCTCGAAGTGAAGACGCGCGCTGGAGAACATCAAAA 214

Qy 133 LeuGlyLysGlyIleAsnGlnAlaSerValGluIleLeuSerPheGlyGluArgPro 152

Db 215 CTAGCCCAAGAGTCTC---AAGCTGCGTATTGAGACCTCGAGTGGCAACAGAGAGATC 271

Qy 153 IleAlaPheGlyThrAsn 158

Db 272 GCCGCTGCCGCCACCAAC 289

RESULT 33

LOCUS BE135571 480 bp mRNA linear EST 21-JUN-2000

DEFINITION u953a03.y1 Bartshead bowel MFLR9 Mus musculus cDNA clone

IMAGE:1546060 5', similar to TR:060817 060817 NASCENT

POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE ;, mRNA sequence.

ACCESSION BE135571

VERSION BE135571.1 GI:8598090

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 480)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 480)

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

Contact: Maria M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:951408

Seq primer: -40RP from Gdbco
High quality sequence stop: 474.

FEATURES

Source

1. 480

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone_lib="Bartstead bowel MLRB9"

/tissue_type="bowel"

/dev_stage="8 weeks"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5];

TGTTACGAATCTGAACTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT

3/1; double-stranded cDNA was ligated to Eco RI adaptors

[AATTCGATCCTTG], digested with Not I and cloned into the

Not I and Eco RI sites of the modified pT73 vector.

Source irradiated bowel harvested 72 hours after

irradiation (1400 Gys). Library constructed by Bob

Bartstead.

BASE COUNT 149 a 118 c 125 g 88 t

Alignment Scores:

Score: 6.98 Length: 480
Percent Similarity: 79.50 Matches: 40
Best Local Similarity: 35.03% Conservative: 15
Query Match: 25.48% Mismatches: 55
9.23% Indels: 47
Gaps: 5

US-09-674-779b-2 (1-172) x BE135571 (1-480)

25 AaLysSerThrSerglnValMetValAlaProAlaProThrGlyThrGlyVal 44

1 AATTCGTCACATCTTGTCGCGATCTCCGCAACAAATCCCGG 48

45 IleTyrThrGlyValAlaProLeuValAspAsnAspGlnThrValLysAlaLeuAlaSer 64

49 -----GAAGCCACAGAAACCGCTCCGCTACAGACAGCAG 81

65 LysLeuProSerLeuValTyr-----PheAspPheAspSerAspGlnIle 79

82 GAGTTCGCCACAGCCCTCAGGCTGACAGATCGGAAACAGAGCTGACAGTGAAGTCA 141

80 LysProGlnAlaAlaAlaIleLeuAspGlnGlnAlaIlePheLeuThrThrAsnGlnThr 99

142 GTACCAGAG-----CTCGAGGACACAGACTCCACACAGAGCGCCACGACGAA 189

100 AlaArgValLeuValAlaGlyHisThrAsp----- 109

190 GCCCAGCTGGCAGCCGACAGACAGATCGATGAAGAACCCTGTAGTAAGCCACAGAGT 249

110 -----GluArgGlySerArgGlyLysArgLysMetSerLeuGlyLysArgAlaValAla 127

250 CGAAGTGAAGAAAGCAAGCAAGCAAGCTATGTCACAACTGGGCTCTTCGACAGTTACAGGG 309

128 ValArgAsnTyrLeuLeuGlyLysGlyLysLeuAsn----- 138

310 GTTACGAGAGTCACTATCCGAAATCTAAATATCTCTTGTGATCACAACAAACCGAT 369

139 -----GlnAlaSerValGlnIleIleSerPheGlyGlnGlyArg 151

370 GTCTAAGAGAGCCAGCTTCAGACACCTACATAGTGTGGGGAAGCCAG 420

DB

QY

RESULT 34

CNS075LQ

LOCUS

CNS075LQ 919 bp DNA linear GSS 08-JUL-2001

clone BA0AB036F06 of library BA0AB from strain CLIB 210 of

Kluyveromyces lactis, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

1152876 2 (bases 1 to 919)

Boletín-Pukhara,M., Toffano-Nioche,C., Attiguenave,F.,

Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,

Robert,C., Termier,M., Wincker,P. and Weslowski-Louvel,M.

Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

FEMS Lett. 487 (1), 66-70 (2000)

2 (bases 1 to 919)

Boletín-Pukhara,M., Toffano-Nioche,C., Attiguenave,F.,

Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,

Robert,C., Termier,M., Wincker,P. and Weslowski-Louvel,M.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEMS Lett. 487 (1), 3-12 (2000)

11152876 3 (bases 1 to 919)

Boletín-Pukhara,M., Toffano-Nioche,C., Attiguenave,F.,

Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,

Robert,C., Termier,M., Wincker,P. and Weslowski-Louvel,M.

Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

FEMS Lett. 487 (1), 66-70 (2000)

20584721 3 (bases 1 to 919)

Boletín-Pukhara,M., Toffano-Nioche,C., Attiguenave,F.,

Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,

Robert,C., Termier,M., Wincker,P. and Weslowski-Louvel,M.

Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

FEMS Lett. 487 (1), 66-70 (2000)

20584721 3 (bases 1 to 919)

Boletín-Pukhara,M., Toffano-Nioche,C., Attiguenave,F.,

Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,

Robert,C., Termier,M., Wincker,P. and Weslowski-Louvel,M.

Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

FEMS Lett. 487 (1), 66-70 (2000)

20584721 3 (bases 1 to 919)

Boletín-Pukhara,M., Toffano-Nioche,C., Attiguenave,F.,

Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,

Robert,C., Termier,M., Wincker,P. and Weslowski-Louvel,M.

Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

FEMS Lett. 487 (1), 66-70 (2000)

20584721 3 (bases 1 to 919)

Boletín-Pukhara,M., Toffano-Nioche,C., Attiguenave,F.,

Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,

Robert,C., Termier,M., Wincker,P. and Weslowski-Louvel,M.

Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

FEMS Lett. 487 (1), 66-70 (2000)

20584721 3 (bases 1 to 919)

Boletín-Pukhara,M., Toffano-Nioche,C., Attiguenave,F.,

Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,

Robert,C., Termier,M., Wincker,P. and Weslowski-Louvel,M.

Genomic exploration of the hemiascomycetous yeasts: 11.

Kluyveromyces lactis

FEATURES

Source

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-674-779b-2 (1-172) x CNS075LQ (1-919)

QY

DB

QY

18 ThrPheMetThrGlyCysAlaAsnLysSerThrSerglnValMetValAlaProAla 37

250 ACAATTAGCGGTGGGCTAGTAACACTACTACTAGTCTGAT-----GTACATCAACAGAG 303

38 ProThr-----GlyTyrThrGlyValIleTyr 46

```

Db      304 CCAACCTTCATGACTACTGCTACACAAAGTAATGGCAGTACC---ACATATTAT 360
QY      47 ThrGlyValAlaProLeu-----ValAspAsnAspGluThrValLys 60
Db      361 ACCACTACTTCCCTATCGATACACACTGTACCGAGAGTGATGATGACCAACATATCT 420
QY      61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLys 80
Db      421 GATGTAAATATCACACAGCCAACTTTC----- 447
QY      81 ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
Db      448 -----TCCTGACTACTGCTACACAAAGTAACGCTGTACTACCGCTTACACACACC 501
QY      101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlyTyrAsnMetSerLeu 120
Db      502 ACTGTCTCTATCGATACACACTGTACCAAGTGCATCTAATGTAGTACACACTTCAAG 561
QY      121 GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGln--- 139
Db      562 TCGGAGATCGATGAATTTCTACACAGAAACCTCATTTGCTACAGGTACACAAACAAAGAC 621
QY      140 -----AlaSerValGluIleIleSerPheGlyGlu-----GluArgPro 152
Db      622 CAAGTCTCTAGTGTGAAGTAACAACCGCTTCTCAAAATGCTGAAGTCTCATCAGCCGAC 681
QY      153 IleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgArgAlaGluLeuSer 171
Db      682 ATCGCTATCACTCTGATGTGTCAATCGAAAGTGAACACAGACACAAACAAAG 738

```

```

RESULT 35
W20805      460 bp      mRNA      linear      EST 10-SEP-1996
LOCUS      MB91C06.1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
DEFINITION IMAGE:36778 5' similar to PIR:S49326 S49326 Nascent polypeptide
            associated complex alpha chain - human ; mRNA sequence.
ACCESSION  W20805
VERSION     W20805.1 GI:1297704
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 460)
            Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterson,R.
TITLE       The WashU-HMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     Contact: Maria M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:218178
            Seq primer: EMP1mer
            High quality sequence stop: 454.

```

FEATURES

```

source      1. 460
            location/Qualifiers
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:336778"
            /clone_lib="Soares mouse p3NMF19.5"
            /dev_stage="19.5 ope total fetus"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Vector: pT73D (Pharmacia) with a modified

```

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCCGCGATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

```

BASE COUNT      133 a      119 c      122 g      86 t
ORIGIN
Alignment Scores:
Pred. No.:      7.54      Length:      460
Score:          79.00      Matches:      39
Percent Similarity: 35.26%      Conservative: 16
Best Local Similarity: 25.00%      Mismatches: 57
Query Match:      9.18%      Indels:      44
DB:              14      Gaps:      5

```

US-09-674-779b-2 (1-172) x W20805 (1-460)

```

QY      26 LysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIle 45
Db      24 CGTTCTCTTTTCTGCCCGCATCTTGTTCCGTATCTCCACAAATGCCCGT----- 77
QY      46 TyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLys 65
Db      78 -----GAAGCCACAGAAACCGCTCCCTCTACAGACAGAGG 113
QY      66 LeuProSerLeuValTyr-----PheAspPheAspSerAspGluLys 80
Db      114 TTGGCCACAGCCTCAGGCTGAGACAGATCGGAGACAGAGCTGACATGATAGTCAGTA 173
QY      81 ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
Db      174 CCAAG-----CTCGAGAAACAAGACTCCACACAGACGCGCAGACAGAGCC 221
QY      101 ArgValLeuValAlaGlyHisThrAsp----- 109
Db      222 CAGCTGGCAGCCCGCAGCAGAGATCGATGAAGAACTGTATGTAAGCAAGCAGAGCTGA 281
QY      110 -----GluArgGlySerArgGlyTyrAsnMetSerLeuGlyGluArgArgAlaValAlaVal 128
Db      282 AGTGAGAGAGAGGACGAGAAAGGCTATGTCCAAACTGGCTTTCGACAGGTTACAGGGGTT 341
QY      129 ArgAsnTyrLeuLeuGlyLysGlyIleAsn----- 138
Db      342 ACGAGAGTCACTATCCGAAATCTAATAATATCTCTTGTCTATCACAACCAACCGATGTC 401
QY      139 -----GlnAlaSerValGluIleIleSerPheGlyGluGluArg 151
Db      402 TACAAGAGCCCGACCTTCAGACACCTCATATGTGTGGGGAACCAAGA 449

```

RESULT 36
BG261745 1083 bp mRNA linear EST 13-FEB-2001
LOCUS BG261745
DEFINITION 602375532F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4481142 5',
mRNA sequence.
ACCESSION BG261745
VERSION BG261745.1 GI:12771561
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1083)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-r@mail.nih.gov

Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: LLM10315 row: p column: 07
 High quality sequence stop: 641.

FEATURES

Location/Qualifiers
 1. 1083

BASE COUNT 295 a 276 c 285 g 227 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 31.3 Length: 1083
 Score: 79.00 Matches: 30
 Percent Similarity: 44.04% Conservative: 18
 Best Local Similarity: 27.52% Mismatches: 38
 Query Match: 9.18% Indels: 24
 DB: 12 Gaps: 4

US-09-674-779b-2 (1-172) x BG261745 (1-1083)
 QY 55 AsnAspGluThrValValAlaLeuAlaSerLysLeuProSer-----LeuValTyr 71
 Db 485 AGCATGCAATTCACAGAAATACCCAGAGATCTCCCTGTAACAACTTTGCTGTAT 544
 QY 72 PheAspPheAspSerAspGluLeuLysProGln----- 82
 Db 545 TTGGACTTC-CATTCAGATTCATTCATCCCATGAGATTTTTAAGGACCTCCATCAAC 603
 QY 83 -----AlaAlaAlaIleLeuAspGluGlnAlaGlnPhe 93
 Db 604 TGAGAGAGGTTAACTGTGCAAAAAACGGCATCGAGTTATTTGGACGACGATTGTC--TTC 660
 QY 94 LeuThrThrAsnGlnThrAlaArgValLeu--ValAlaGlyHisThrAspGluArgGly 112
 Db 661 CAAGGGGTAGAGAAACCCCTGGAGAACCTGGACTGTGTGGCCACCGGATTCACAGGGGT 720
 QY 113 SerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyrLeu 132
 Db 721 GCCCAAAAATGCTCTCAATCTCTGGAAGGCAAGGCTGATTAACCCACGCG 780
 QY 133 LeuGlyLysGlyLysGlnAlaSer 141
 Db 781 GTTGTGGGGGGGTCTCCCAAGTTTCC 807

RESULT 37
 BE261626 503 bp mRNA linear EST 26-OCT-2000
 LOCUS 60119316F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3502299 5',
 DEFINITION mRNA sequence.
 ACCESSION BE261626
 VERSION BE261626.1 GI:9133944
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 503)
 AUTHORS NIH-MGC <http://mgs.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at: image.llnl.gov
 Plate: LLM173 row: o column: 04
 High quality sequence start: 3
 High quality sequence stop: 502.

FEATURES

Location/Qualifiers
 1. 503

BASE COUNT 105 a 134 c 141 g 123 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 10.1 Length: 503
 Score: 78.50 Matches: 26
 Percent Similarity: 42.45% Conservative: 33
 Best Local Similarity: 18.71% Mismatches: 61
 Query Match: 9.12% Indels: 19
 DB: 10 Gaps: 5

US-09-674-779b-2 (1-172) x BE261626 (1-503)
 QY 23 CysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProHngIYrThr 42
 Db 55 TGCTCTCGCACACACCCACTTCATTGTGGCAACCAACCGACCCCGAATTCTCC 114
 QY 43 GlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeu 62
 Db 115 GCMAATTTCTGTCCGGCTAGGTCACGATTTGAGGTGATGATTCGTGTGTTT 174
 QY 63 AlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLys----- 80
 Db 175 GAGACCAAC-----GCCATTGCTTACTAT--GTGACCAATGAGGAGCTGGCGGCAAGT 225
 QY 81 ---ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThr 99
 Db 226 ACTCCAGAGCAGACGACGACGCTGGTGCAGTGGTGTGCTGATTCGACCTAAGT 285
 QY 100 AlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlyTyrAsnMetSer 119
 Db 286 GCCCCACGACGATCTGTGTGTGTCCTCCACCTTGGGCAATGACCAACAACAG 345
 QY 120 LeuGlyGlu-----ArgArgAlaValAlaVal----- 128
 Db 346 GCCACTGAGATGAAAG 405
 QY 129 -----ArgAsnTyrLeuLeuGlyLysGlyTyrAsnGlnAlaSerValGluIle 145
 Db 406 AAGACAGAGACTTTCTGTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462

RESULT 38
 D34784 360 bp mRNA linear EST 08-AUG-1994
 LOCUS D34784

DEFINITION	CELK012E3F yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA				
ACCESSION	D34784				
VERSION	D34784.1 GI:526289				
KEYWORDS	EST.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans				
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea 1 (bases 1 to 360) Kohara,Y., Mitsunuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.				
TITLE	Toward an expression map of the C.elegans genome				
JOURNAL	Unpublished (1994)				
COMMENT	Contact: Yuji Kohara Genome Biology Lab. National Institute of Genetics Yata 1111, Mishima, Shizuoka 411, Japan Tel: 81-559-81-6854 Fax: 81-559-81-6855 Email: ykohara@lab.nig.ac.jp.				
FEATURES	Location/Qualifiers				
SOURCE	1..360				
	/organism="Caenorhabditis elegans"				
	/strain="CB1489 him-8(e1489)"				
	/db_xref="taxon:6239"				
	/clone="YK123"				
	/clone_lib="Yuji Kohara unpublished cDNA"				
	/sex="hermaphrodite, male"				
	/tissue_type="whole animal"				
	/dev_stage="varied"				
BASE COUNT	74 a	96 g	76 t	2 others	
ORIGIN	112 c				
Alignment Scores:					
Pred. No.:	6.75	Length:	360		
Score:	78.00	Matches:	39		
Percent Similarity:	41.38%	Conservative:	21		
Best Local Similarity:	26.90%	Mismatches:	43		
Query Match:	9.06%	Indels:	42		
DB:	14	Gaps:	8		
US-09-674-779B-2 (1-172) x D34784 (1-360)					
OY	4	Histl	ecl	unil	ea
Db	34	TACCT	CAAT	TCGAG	CGCCGCGCAGTCACCATGGCGCTCGGATCTACTGTGTGAAG
OY	24	Ala	sn	ys	ser
Db	94	AGCT	CG	CA	TCGACGAGTCG-----GGATCTACTGCTTCAGTCGG
OY	44	Val	l	e	tyr
Db	133	GTC	AC	TAT	TCGGGAATGTCAGGCACCTGCTGATCCGATGCAATCGAATCGGATTTTG
OY	61	Ala	leu	na	la
Db	193	GCG	AG	T	GTTCGCGGACCCGCTCTCGAACC-----
OY	81	Pro	g	l	na
Db	211	AGC	GC	T	ACCTACCTCGAAGCAGACGATCCAGGCTTCTCTGTTCC-----GAGAC
OY	100	Ala	arg	val	leu
Db	259	GCG	GC	GC	ACCTCTACCCAGGGT-----GTTCGCGGACCCGCTCTCGAACC-----
OY	120	Leu	g	l	yl
Db	307	-----	-----	-----	-----
OY	140	Ala	ser	val	glu
Db	-----	-----	-----	-----	-----

```

DB          340  TCGATGCTTCATC 354
RESULT 39
A0954996/c 470 bp  DNA  linear  GSS 27-JAN-2000
LOCUS
DEFINITION
nbe0076m14f CUG1 Rice BAC library (EcoRI) Oryza sativa genomic
clone OSJNB0076m14f, DNA sequence.
ACCESSION
A0954996
VERSION
A0954996.1 GI:6778262
KEYWORDS
GSS.
SOURCE
Oryza sativa.
ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 470)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GMAAACGACGCCAGTG
Class: BAC ends
High quality sequence start: 73
High quality sequence stop: 332.
Location/Qualifiers
1..470
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="OSJNB0076m14f"
/clone_lib="CUG1 Rice BAC Library (EcoRI)"
/lisse_type="leaf"
/lab_host="E. coli DH10b"
/note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
```

